

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: November 12, 2002, 14:47:29 ; Search time 2236 Seconds  
(without alignments)  
17303.675 Million cell updates/sec

Title: US-08-961-083-55

Perfect score: 2389  
Sequence: 1 TTCTTACGAGTTGGACTGT.....TAAGTAAGCAAAAATAAAC 2389

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_estbun:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70.8	3.0	906	17	BH153606 ENTTS83TF
2	69.8	2.9	890	17	BH146886 ENTPK48TF
3	67.8	2.8	180	17	AF325819 AF325819
4	67.8	2.8	908	17	A2548467 A2548467
5	64.8	2.7	843	17	A2551618 A2551618
6	63.8	2.7	877	17	A2531291 ENTBO34TR

7	62.2	2.6	912	17	A2551092
C	60.8	2.5	849	17	A2546009
8	59.2	2.5	931	17	BH160272
C	58.6	2.5	1101	17	CNS00396
10	57.2	2.4	900	17	A2549980
C	56.8	2.4	905	17	A2550256
12	56.8	2.4	905	17	A2550256
13	56.6	2.4	645	13	BH165350
14	56.6	2.4	747	13	BH162732
C	55.2	2.3	816	17	A2535744
15	55.2	2.3	647	17	BH163120
16	55	2.3	671	13	BH160252
17	55	2.3	694	13	BH168242
18	55	2.3	717	13	BH160500
19	55	2.3	756	13	BH162492
20	55	2.3	794	13	BH159906
21	55	2.3	795	17	A2528485
22	55	2.3	823	17	A2676218
C	55	2.3	843	17	BH139532
23	55	2.3	890	17	A2530768
24	52.8	2.2	677	13	BH161314
25	52.4	2.2	725	13	BH169486
26	52.4	2.2	866	17	A2527885
27	52.4	2.2	976	17	BH149983
28	52.4	2.2	945	17	CNS008CA
29	52.2	2.2	435	13	BH169710
30	51.8	2.2	880	17	A2529191
C	51.8	2.2	853	17	CNS023KH
32	51.6	2.2	500	9	AU088119
C	51	2.1	726	17	A2196050
34	50.8	2.1	1007	17	CNS06X9S
C	50.2	2.1	445	13	BH167469
35	50	2.1	707	13	BH164483
36	49.8	2.1	598	13	BH170666
C	49.6	2.1	605	13	BH163520
37	49.6	2.1	629	13	B073162
38	49.6	2.1	687	13	BH169665
39	49.4	2.1	1101	17	CNS0022U
40	48.8	2.0	605	13	BH162501
41	48.6	2.0	891	17	A2683582
42	48.6	2.0			
43	48.6	2.0			
44	48.6	2.0			
45	48.4	2.0			

#### ALIGNMENTS

RESULT 1  
LOCUS BH153606/c 906 bp DNA linear GSS 24-SEP-2001  
DEFINITION ENTTS83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
ACCESSION BH153606  
VERSION BH153606.1 GI:15725323  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 906) Eukaryota; Entamoebidae; Entamoeba.  
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library (2001)  
COMMENT Unpublished (2001)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@igf.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Forward  
Class: Shotgun  
High quality sequence start: 15  
High quality sequence stop: 733.





1645 TCATATGGCCATAGTCATGATG---GAAAGATAGCCTTCTGATTAAGAAAGT 1701  
153 TGAAGACGATGAAGACGACCAATATGATTAAGAGATGATGATGAAGAAGAGA 212  
1702 TGCAGCTCAAGCCTATTAAGAAAAAGGTATCTACCTCATCTCCAGACGCAGATGT 1761  
213 TGATGATGAAGATGATGAAGACGATGAAGACGAAATATGAATTAAGATGATGATGA 272  
1762 TAAGCAAAATCCAACTGGAGATAGTGCAGACTATTACAAATCGTGTGAAGGGGAAA 1821  
273 TGATGAAGAAGAGATGATGATGAAGAAGAGATGATGAAGATGATGATGAAGA 332  
1822 ACAATTCCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1881  
333 AGAAGATGAAGAAGATGATGAAGAAGATGATGAAGATGATGAAGAAGATGA 392  
1882 TAATTTGATTAATCTCCATTAAGATCATTAACATTAATTAATTAATTTGCTTGGTTGATGA 1941  
393 TGATGAAGATGATGAAGACGATGAAGACGAAATATGAATTAAGATGATGATGAAGA 452  
1942 TCACACATTAAGATCCCAATGGCTATACCTTGAAAGATTTGTTGCCGATTAAGTA 2001  
453 AGAAGAAGATGATGATGAAGAAGAAGATGATGATGAAGATGATGAAGAAGAAGA 512  
2002 CTACGTAGAAGACCCCTGACGACGATCCACATTTCTAATGATGATGGGGCAATGCCAGTGA 2061  
513 TGAAGATGATGATGAAGAAGATGATGAAGATGATGAAGAAGATGATGATGAAGA 572  
2062 GCATGTTTGAAGCAAGAAAGACCAAGTGAAGATCCAAATTAAGACTTCAAGCGCATGA 2121  
573 AGAAGATGAAGAAGAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGA 632  
2122 AGAGCCAGTGAAGAAACCTCTGACGACGAGAGTCCCTCACTGATGAGCTGAAAAAGT 2181  
633 AGACGACGATGATGAAGATTAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGA 692  
2182 AGAAGCCCACTCAAGAAAGCAGAAAGTTTGTTCGGAAGTAAAGGATCTTCACTGATGA 2241  
693 AGATGATGAAGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGA 752  
2242 AGCCAAATGCACAGAAACTCTAGCTGTTTACGAATTA 2280  
753 TGAAGAAGATGAAGATGATGATGATGAAGAAGATGA 791

RESULT 5  
LOCUS A2551618/c 843 bp DNA linear GSS 14-NOV-2000  
DEFINITION ENT54TFR Entamoeba histolytica sheared DNA Entamoeba histolytica  
ACCESSION A2551618  
VERSION A2551618.1 GI:11176919  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica.  
REFERENCE 1 (bases 1 to 843)  
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
HMI:IMSS sheared DNA library  
JOURNAL Unpublished (2000)  
COMMENT Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@igr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library  
Seq primer: M13-Reverse  
Class: shotgun

High quality sequence start: 39  
High quality sequence stop: 838.  
Location/Qualifiers  
1..843  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica sheared DNA"  
/note="Vector: PHOS1. Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.) The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + l method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 155 a 266 c 18 g 404 t  
ORIGIN

Query Match 2.7%; Score 64.8; DB 17; Length 843;  
Best Local Similarity 44.1%; Pred. No. 3.4e-05;  
Matches 319; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

QY 1521 AAATTGATTAAGTGAAGACGAGTTCGTTCTCAATTAAGTGAATTAAGTGAACGT 1580  
DB 753 AAGAGAGTGTGATGAAGAAGATGATGAAGATGATGAAGATGATGAAGATGA 694  
QY 1581 CAGATGTTACAT---TTTGATGAACATGATTAATCAAGTGAAGAGATGATGATG 1637  
DB 693 AAGATGATGATGAAGAAGAGATGATGAAGATGATGAAGATGATGAAGATGATGA 634  
QY 1638 TAAGCCCTCATATGAGCCATGATGATGATGATGATGATGATGATGATGATGATG 1697  
DB 633 AGCAATATGATTAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGA 574  
QY 1698 AAGTTGACGCTCAAGCCTTACTTAAGAAAGATGATGATGATGATGATGATGATG 1757  
DB 573 AAGATGATGATGAAGATGATGATGAAGAAGATGAAGATGAAGATGAAGATGAAG 514  
QY 1758 ATGTTAAGCAATTCACACTGAGATGATGATGATGATGATGATGATGATGATGATG 1817  
DB 513 AAGATGATGATGAAGAAGAGATGATGATGAAGATGATGATGAAGATGATGAAG 454  
QY 1818 AAAAAGAAATTCACCTGCTTCATATATGATGATGATGATGATGATGATGATGAT 1877  
DB 453 AATATGAATTAAGAAAGATGATGAAGAAGATGAAGATGATGAAGAAGATGAAG 394  
QY 1878 ACGTAAATTTGATTAATCTCATTAAGATCATTAATTAATTAATTTGCTGTTGG 1937  
DB 393 ATGATGAAGATGATGAAGAAGAGATGATGAAGAAGATGATGAAGAAGATGAAG 334  
QY 1938 ATGATCAACATTCAAAGCTCCAAATGGCTATACCTTGAAGATTTGTTGGCAGATTA 1997  
DB 333 ATGATGAAGAGATGATGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG 274  
QY 1998 AGTACTACGTAGAAGACCCCTGACGACGATCCATTAATGATGATGATGATGATG 2057  
DB 273 ATGATGAAGAAGAGATGATGAAGAAGATGATGAAGATGATGAAGAAGATGAAG 214  
QY 2058 GTGAGCATGTTGAGCAAGAAAGACACAGTGAAGATCCAAATTAAGATTCGAAGCG 2117  
DB 213 AAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGATGAAG 154  
QY 2118 ATGAAGACGATGAAGAAAGACCTGCTGAGCAGCAAAATCCCTCAAGTGAAGACTG 2177  
DB 153 ATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGAAG 94  
QY 2178 AAGTAGAACCCCAACTCAAGAAGAGTTTGTTCGAAAGTAAAGATTTAGTC 2237



[illegible][illegible]

RESULT	7
AZ551092	
LOCUS	AZ551092
DEFINITION	AZ551092 912 bp DNA linear GSS 14-NOV-2000
ACCESSION	ENFJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
VERSION	AZ551092
KEYWORDS	AZ551092.1 GI:1176393 GSS.
SOURCE	Entamoeba histolytica.
ORGANISM	Entamoeba histolytica
REFERENCE	Eukaryota; Entameobidae; Entamoeba.
AUTHORS	1 (bases 1 to 912)
TITLE	Lofthus,B., Van Aken,S. and Fraser,C. Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Brendan J Lofthus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0208 Fax: 301 838 3543 Email: b.lofthus@tigr.org Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
FEATURES	Seq primer: M13-Forward
source	Class: Shotgun High quality sequence start: 17 High quality sequence stop: 861. Location/Qualifiers 1..912 /organism="Entamoeba histolytica" /strain="HMI:IMSS" /db_xref="taxon:5759"



```

Db      442 GATGATGAATTTGAATTAGAGATGAGAGATGAAGATGAAGATGATGATGAT 383
Qy      1550 ATTGCTCAATTTAGCGTATAGATATACAGTCAGATGCTTACATTTTGTATCAATGAT 1609
Db      382 GATGATGAAGATGAAGATGAAGATGATGATGAAGATGATGAAGATGATGAAGATGAT 323
Qy      1610 ATATATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1669
Db      322 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 263
Qy      1670 GGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1729
Db      262 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 203
Qy      1730 GGATCTTACCTCCATCTCCACAGCAGATGTTAAAGCAATCCACATGAGATGATGAT 1789
Db      202 GAATTAAGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGAAGATGA 143
Qy      1790 GCAGCTATTACATCGTGTGAAGGCGGAAAA 1822
Db      142 GAAGAAGAAGATGATGATGATGAAGAGAA 110

```

```

RESULT 9
LOCUS   BH160272
DEFINITION  ENTQV49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
          genomic, DNA sequence.
ACCESSION   BH160272
VERSION     BH160272.1 GI:15733710
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica.
REFERENCE   1 (bases 1 to 931)
AUTHORS     Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
            HMI:IMSS sheared DNA library (2001)
            Unpublished (2001)
JOURNAL     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@igf.org
            Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 8
            High quality sequence stop: 677.
FEATURES
    source
        1..931
            location/Qualifiers
                /organism="Entamoeba histolytica"
                /strain="HMI:IMSS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: PHOSI; Site: 1; Bst I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                tight size distribution (~2 kb). The v + i method used for
                the library construction is described in detail in Smith,
                H.O. and Venter, J.C. (Making small insert libraries for
                whole genome shotgun sequencing projects. In Genome
                Sequencing: A Practical Approach, eds. M. Vaubin and B.
                Barrell, Oxford University Press, 1999)."
```

BASE COUNT

```

402 a
51 c
279 g
199 t

```

## ORIGIN

Query Match 2.5%; Score 59.2; DB 17; Length 931;  
 Best Local Similarity 43.3%; Pred. No. 0.00086;  
 Matches 277; Conservative 0; Mismatches 363; Indels 0; Gaps 0;

```

Qy      1583 GATGCTTACATTTTGTATGATCAATGATATATATCACTGATGAAGAGATGATGATTAACG 1642
Db      17 GATGATGATGAAGAGAGAGAGATGATGATGAAGATGATGAAGAGATGATGAAGAGAG 76
Qy      1643 CCTCATATGGGCCATAGTCCTGATGGAAGAAGATGACCTTTCGTATGAAGAAAAAGTT 1702
Db      77 TATGATTAAGAAGATGATGATGATGATGATGAAGAAGATGATGATGAAGAAAAAGAT 136
Qy      1703 GCAGCTCAAGCCTATACATAAAGAAAAAGTATCCATCCCTCCATCCGACAGCCAGATGT 1762
Db      137 GATGATGAAGATGATGATGAAGAGAGATGAAGAGATGATGAAGAGATGATGAAGAT 196
Qy      1763 AAAGCAATCCAACTGAGATAGTCAGCAGCAGCTATTTTCAATCGTGTGAAGGGAAAAA 1822
Db      197 GATGATGAAGAGAGAGAGAGATGATGATGAAGATGATGAAGAGATGATGAAGAGAGAT 256
Qy      1823 CGAATTCACCTGCTGCACTTCCATATATGTTGACATACACTTGAGGTTAAAAAGCT 1882
Db      257 GAATTAAGAAGATGATGATGAAGAGAGAGATGATGATGAAGAGAGAGATGATGAT 316
Qy      1883 AATTTGATATATCCCATATAGATCATTTCCATATATTAATTTGCTGTTGATGAT 1942
Db      317 GAAGATGATGATGAAGAGAGATGATGATGAAGAGAGATGATGAAGAGAGATGATGAT 376
Qy      1943 CACACATACAAAGCTCCAAATGGCTATACCTTGAAGATTTGTTGCGAGATTAAGTAC 2002
Db      377 GAAGAAGAAGATGATGATGAAGAGAGATGAAGAGATGAAGAGAGATGATGAAGAGAT 436
Qy      2003 TACGTAGAACCACCTGACGACAGCTCCACATTTATGATGATGGGCAATCCAGTGA 2062
Db      437 GAAGAAGAAGAGATGATGATGAAGAGAGATGATGATGAAGAGATGATGAAGAGAT 496
Qy      2063 CATGTGTTAGGCAAGAGAGAGAGAGATGATGAAGATGAAGATGAAGATGAAGATGA 2122
Db      497 GATGAAGAAGATGATGATGAAGAGAGATGATGAAGAGAGATGATGATGATGAT 556
Qy      2123 GAGCCGTTAGAGAAACACCTGCTGAGCCAGCAAGTCCCTCAAGTAGAGATGAAAAAGTA 2182
Db      557 GAAGAAGAAGATGATGATGAAGAGATGAAGAGATGAAGAGATGAAGAGATGAAGATGAT 616
Qy      2183 GAGCCCAACTCAAGAAGACGAGAGATTTTCTTGGCAAG 2222
Db      617 GAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAG 656

```

```

RESULT 10
LOCUS     CNS00396G
DEFINITION  Drosophila melanogaster genome survey sequence TE73 end of BAC #
            BACR08K10 of RPCT-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   CNS00396G
VERSION     AL063921.1 GI:4941778
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster.
ORGANISM    Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1101)
AUTHORS     Genoscope.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 Evry cedex - FRANCE (E-mail : sequefgenoscope.cns.fr
            - Web : www.genoscope.cns.fr)
            Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
```







[illegible]



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 14:47:29 ; Search time 71 Seconds  
(without alignments)  
10319.025 Million cell updates/sec

Title: US-08-961-083-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGGACTGT.....TAAGTAGAATAAATAAAC 2389

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCtus\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2388	100.0	2389	US-08-961-083-55	Sequence 55, Appl
2	2388	100.0	8195	US-08-961-527-94	Sequence 94, Appl
3	987.6	41.3	2290	US-08-961-083-65	Sequence 65, Appl
4	653.8	27.4	2359	US-08-961-527-243	Sequence 243, App
5	385.4	16.1	1342	US-08-961-083-181	Sequence 181, App
6	385.4	16.1	6867	US-08-961-527-192	Sequence 192, App
7	381.4	16.0	973	US-08-961-527-355	Sequence 355, App
8	243.4	10.2	1684	US-08-961-527-258	Sequence 258, App
9	109	4.6	841	US-08-743-637B-34	Sequence 34, Appl
10	109	4.6	841	US-08-526-840B-34	Sequence 34, Appl
11	57.4	2.4	7218	US-08-232-463-14	Sequence 14, Appl
12	51.6	2.2	11091	US-08-134-001C-2243	Sequence 2243, Ap
13	45.2	1.9	2277	US-08-676-967-2	Sequence 2, Appl1
14	45.2	1.9	2277	US-08-676-974-2	Sequence 2, Appl1
15	45.2	1.9	2277	US-09-098-487-2	Sequence 2, Appl1
16	41.4	1.7	5361	US-08-973-462-2	Sequence 2, Appl1
17	41.4	1.7	6152	US-08-973-462-1	Sequence 1, Appl1
18	41.2	1.7	696	US-08-461-687-193	Sequence 193, App
19	41.2	1.7	699	US-08-461-687-191	Sequence 191, App
20	41.2	1.7	717	US-08-461-697-189	Sequence 189, App
21	41.2	1.7	774	US-09-461-697-187	Sequence 187, App
22	41.2	1.7	819	US-09-461-697-185	Sequence 185, App
23	41.2	1.7	1669	US-08-461-687-184	Sequence 184, App
24	40.6	1.7	1704	US-08-134-001C-925	Sequence 925, App
25	40	1.7	30549	US-09-134-001C-322	Sequence 322, App
26	39.8	1.7	297	US-09-134-001C-1027	Sequence 1027, Ap
27	39.8	1.7	1716	US-09-134-001C-1028	Sequence 1028, Ap

28	39	1.6	1839	US-09-134-001C-1191	Sequence 1191, Ap
29	38.8	1.6	1856	US-08-360-606B-29	Sequence 29, Appl
30	38.8	1.6	2150	US-08-861-464-13	Sequence 13, Appl
31	38.8	1.6	2150	US-08-396-001-13	Sequence 13, Appl
32	38.8	1.6	2150	US-09-323-433A-13	Sequence 13, Appl
33	38.6	1.6	6744	US-08-119-125A-2	Sequence 2, Appl1
34	38.6	1.6	8920	US-08-446-855A-1	Sequence 1, Appl1
35	38.6	1.6	8920	US-09-150-741-1	Sequence 1, Appl1
36	38.4	1.6	5319	US-08-169-927-1	Sequence 1, Appl1
37	38.4	1.6	6846	US-08-961-527-198	Sequence 198, App
38	38	1.6	3280	US-08-259-000-4	Sequence 4, Appl
39	38	1.6	3280	US-08-729-767-6	Sequence 6, Appl1
40	37.6	1.6	438	US-09-134-001C-1064	Sequence 1064, Ap
41	37.4	1.6	3294	US-08-923-992A-7	Sequence 7, Appl1
42	37.4	1.6	26385	US-08-961-527-3	Sequence 3, Appl1
43	37.2	1.6	1079	US-07-781-355-1	Sequence 1, Appl1
44	37.2	1.6	2397	US-09-134-001C-2080	Sequence 2080, Ap
45	37	1.5	1103	US-09-122-400B-20	Sequence 20, Appl

## ALIGNMENTS

RESULT 1  
US-08-961-083-55  
Sequence 55, Application US/08961083  
Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Chol et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESSES:

ADDRESSER: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 2389 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-083-55

Query Match 100.0%; Score 2388; DB 3; Length 2389;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2389; Conservative 0; Mismatches 0; Indels 0;

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAATAATCGTTTCTTA 60  
Db 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGTTAAGGAAATAATCGTTTCTTA 60

OY	61	TAGAGATGGAANAACAGCGACGGAAAAACGGAGATTGACTCCGTGAGTACGTATAGCA	120
Db	61	TATGATGGAANAACAGCGACGGAAAAACGGAGATTGACTCCGTGAGTACGTATAGCA	120
OY	121	GGGGAAGGAATCAATGCTGAGCAAAATCGCATCAAGATACAGACCAAGGCTATGTAC	180
Db	121	GGGGAAGGAATCAATGCTGAGCAAAATCGCATCAAGATACAGACCAAGGCTATGTAC	180
OY	181	TTACATGCGCGACCACTATCATTTATTCAAATGGTAAGGTTCCCTATGACGCTATCATG	240
Db	181	TTACATGCGCGACCACTATCATTTATTCAAATGGTAAGGTTCCCTATGACGCTATCATG	240
OY	241	TGAAGATTTACTCATGAAAGATCCAAACTTAAGCTATAAAGATGAGATATTGTTAAAGA	300
Db	241	TGAAGATTTACTCATGAAAGATCCAAACTTAAGCTATAAAGATGAGATATTGTTAAAGA	300
OY	301	GGTCAAGGGTGAATATTATTCAAGGTAGATGGAATAATCTATGTTACTTAAAGATGC	360
Db	301	GGTCAAGGGTGAATATTATTCAAGGTAGATGGAATAATCTATGTTACTTAAAGATGC	360
OY	361	TGCCACGCGGATTAACGTCCGTACAAAAGAGAAATCAATCGACAAAAACAAGACATAG	420
Db	361	TGCCACGCGGATTAACGTCCGTACAAAAGAGAAATCAATCGACAAAAACAAGACATAG	420
OY	421	TCAACATCTGTAGAGTGGAACTCCAGAAACGATGGTGTCTGTGGCTTGGCACGTTGCCA	480
Db	421	TCAACATCTGTAGAGTGGAACTCCAGAAACGATGGTGTCTGTGGCTTGGCACGTTGCCA	480
OY	481	AGGACGCTATCTACAGATATGTTATATCTTAAATGCTTCTGATATCATGAGAGATAC	540
Db	481	AGGACGCTATCTACAGATATGTTATATCTTAAATGCTTCTGATATCATGAGAGATAC	540
OY	541	TGCTGATGCTTATATCGTTCCCTCATGAGAGATCATTCATTCATTCCTTAAGATGAGTT	600
Db	541	TGCTGATGCTTATATCGTTCCCTCATGAGAGATCATTCATTCATTCCTTAAGATGAGTT	600
OY	601	ATCAGCTAGCGAGTGGCTGCTGCGAGAGCTTCTTATCTGTCGAGAAATCTGTCAA	660
Db	601	ATCAGCTAGCGAGTGGCTGCTGCGAGAGCTTCTTATCTGTCGAGAAATCTGTCAA	660
OY	661	TTCAAGAACCTATGCGCGAACAATATGCGCTATCACTTCAAGACAAACTGGGTACCTTC	720
Db	661	TTCAAGAACCTATGCGCGAACAATATGCGCTATCACTTCAAGACAAACTGGGTACCTTC	720
OY	721	TGTAAAGCAATCAGAGACTCAAAATCTAACAACAGCAACAGCAACACTAACAGTCA	780
Db	721	TGTAAAGCAATCAGAGACTCAAAATCTAACAACAGCAACAGCAACACTAACAGTCA	780
OY	781	AGCAAGTCAAGTATATACATTATATGCTCTTTGAACAGCTCTACAAATCGCTTTAG	840
Db	781	AGCAAGTCAAGTATATACATTATATGCTCTTTGAACAGCTCTACAAATCGCTTTAG	840
OY	841	TCAAGCAATGTAAATCTGATGGCCCTGTGCTTGTGTCAGACAACAAATCACAGTGCAC	900
Db	841	TCAAGCAATGTAAATCTGATGGCCCTGTGCTTGTGTCAGACAACAAATCACAGTGCAC	900
OY	901	AGCTAGAGGTGTGTCAGTGGCACAGGAGATCTTAACCACTTCATCCCTTACTCTCAAT	960
Db	901	AGCTAGAGGTGTGTCAGTGGCACAGGAGATCTTAACCACTTCATCCCTTACTCTCAAT	960
OY	961	GCTGGAATTGGAGAGAGAAATCGCTGATATTTCGCCCTTGTTATCTGTTCAACCATG	1020
Db	961	GCTGGAATTGGAGAGAGAAATCGCTGATATTTCGCCCTTGTTATCTGTTCAACCATG	1020
OY	1021	GGTACCGAATTCAAGGCCCAACCAACCAAGTCCACACCGCATCCGGAACTAGTCCAG	1080
Db	1021	GGTACCGAATTCAAGGCCCAACCAACCAAGTCCACACCGCATCCGGAACTAGTCCAG	1080
OY	1081	CCCGCAACTGCAACAATCTTAAATATAGCTCAAAATCTTCTTTGGTTAGTCACTGGT	1140
Db	1081	CCCGCAACTGCAACAATCTTAAATATAGCTCAAAATCTTCTTTGGTTAGTCACTGGT	1140
OY	1141	ACGAAAGTTGGGAGAGATATGTTATTCGAGAAAAAGGCACTCTCTGTTATGCTTTGGC	1200

Db	1141	ACGAAAGTTGGGAGAGATATGATTCCAGAGAAAGGGCATCTCTCGTATGTTTTC	1200
QY	1201	GAAAGATTACCATCTGAAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAGAG	1260
Db	1201	GAAAGATTACCATCTGAAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAGAG	1260
QY	1261	TGTTTCACACATTTAACTCTGTAAGAAAGAAATGTTGCTCTCGTGACCAAGATTTTA	1320
Db	1261	TGTTTCACACATTTAACTCTGTAAGAAAGAAATGTTGCTCTCGTGACCAAGATTTTA	1320
QY	1321	TGATTAACCATATATCTGTGTTAACTGAGGCTCATTAAGCTTGTTGNAATTAAGGGTCG	1380
Db	1321	TGATTAACCATATATCTGTGTTAACTGAGGCTCATTAAGCTTGTTGNAATTAAGGGTCG	1380
QY	1381	TAATTCGATTTCCAGCCCTTAGACAAATATATAGAACCTTGAATGATGCATCTAA	1440
Db	1381	TAATTCGATTTCCAGCCCTTAGACAAATATATAGAACCTTGAATGATGCATCTAA	1440
QY	1441	TAAAGAAAAATGTTAGATGATTTATTGGCATTCCTAGACCAATTTCCATCCAGAGC	1500
Db	1441	TAAAGAAAAATGTTAGATGATTTATTGGCATTCCTAGACCAATTTCCATCCAGAGC	1500
QY	1501	ACTGGCAACCAATCTCAATTAAGTATACTGAAAGCAAGTTCGATATGCTCATTT	1560
Db	1501	ACTGGCAACCAATCTCAATTAAGTATACTGAAAGCAAGTTCGATATGCTCATTT	1560
QY	1561	AGCTGATTAAGTATCAACGTCAGATGGTTACATTTTGTGTAACATGATATATCAGTA	1620
Db	1561	AGCTGATTAAGTATCAACGTCAGATGGTTACATTTTGTGTAACATGATATATCAGTA	1620
QY	1621	TGAAGAGATCATATGTATTAACGCTCATATGGCCATGTCATCGATGGTGAAGATAG	1680
Db	1621	TGAAGAGATCATATGTATTAACGCTCATATGGCCATGTCATCGATGGTGAAGATAG	1680
QY	1681	CCTTTCTGATTAAGAAAAAGTTGCAGCTCAAGCTTACTTAAGAAAAAGTATCTTACC	1740
Db	1681	CCTTTCTGATTAAGAAAAAGTTGCAGCTCAAGCTTACTTAAGAAAAAGTATCTTACC	1740
QY	1741	TCCATCTCCACAGCCAGATGTTTAAAGCAATCCCACTGGAGATAGTGCAGCAGTATTTA	1800
Db	1741	TCCATCTCCACAGCCAGATGTTTAAAGCAATCCCACTGGAGATAGTGCAGCAGTATTTA	1800
QY	1801	CAATCGTGTGAAGGGGAAAAAGAAATTCOCATCGTTGACCTTCATATATGTGTGACA	1860
Db	1801	CAATCGTGTGAAGGGGAAAAAGAAATTCOCATCGTTGACCTTCATATATGTGTGACA	1860
QY	1861	TACAGTTAGAGTTAAAAACGGTATTTGATTAATTCCTCTAAAGGATCATTTACCATTAAT	1920
Db	1861	TACAGTTAGAGTTAAAAACGGTATTTGATTAATTCCTCTAAAGGATCATTTACCATTAAT	1920
QY	1921	TAAATTTGCTGGTTTGTATGATCAGCATCAACATCAAAAGTCCAAATGGCTTATACCTTGAGTA	1980
Db	1921	TAAATTTGCTGGTTTGTATGATCAGCATCAACATCAAAAGTCCAAATGGCTTATACCTTGAGTA	1980
QY	1981	TTTGTGTCGACGATTAAGTACTAGTAGAACCCCTGACGACGCTCCACATTTCAATGA	2040
Db	1981	TTTGTGTCGACGATTAAGTACTAGTAGAACCCCTGACGACGCTCCACATTTCAATGA	2040
QY	2041	TGGATGGGCAATGCCAGTAGCATGTGTTAGCAAGAAAGCCACAGTGAAGATCCAA	2100
Db	2041	TGGATGGGCAATGCCAGTAGCATGTGTTAGCAAGAAAGCCACAGTGAAGATCCAA	2100
QY	2101	TAAAGACTTCAAACGGGATGAAGAGCCAGTAGAGGAACACTGCTGAGCCAGAAATGCC	2160
Db	2101	TAAAGACTTCAAACGGGATGAAGAGCCAGTAGAGGAACACTGCTGAGCCAGAAATGCC	2160
QY	2161	TCAAGTAGAGCATGAAGAAATAGAACCCCACTCAAAAGAGAGAGATTTGCTTGCGAA	2220
Db	2161	TCAAGTAGAGCATGAAGAAATAGAACCCCACTCAAAAGAGAGAGATTTGCTTGCGAA	2220
QY	2221	AGTAACGATTTAGTCTGAAGCCATGCAACAGAAACTTAGCTGATTAACGAATTA	2280

Db 2221 AGTAAGGATTCATAGCTGAAAGCCAAATGCAACGAAGAACTCTAGCTGTAGCAAAATAA 2280  
QY 2281 TTTACGCTTCAATATTTGATATACATAGTATCATGCGAGAGCAAAAAATTTACTTGC 2340  
Db 2281 TTTACGCTTCAATATTTGATATACATAGTATCATGCGAGAGCAAAAAATTTACTTGC 2340  
QY 2341 GTTGTAAAGGAAGTAATCTTCATCTGTAGTAAGGAAAAATAAAC 2389  
Db 2341 GTTGTAAAGGAAGTAATCTTCATCTGTAGTAAGGAAAAATAAAC 2389

## RESULT 2

US-08-961-527-94  
; Sequence 94, Application US/08961527  
; Patent No. 6420135  
GENERAL INFORMATION:  
; APPLICANT: Charles Kunsch  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 8195 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: Linear  
; US-08-961-527-94

Query Match 100.0%; Score 2388; DB 4; Length 8195;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTTCTTACGAGTTGGAGCTGTATCAAGCTAGACGGTTAAAGAAAAATATCGTGTTCCTTA 60  
Db 3053 TTTCTTACGAGTTGGAGCTGTATCAAGCTAGACGGTTAAAGAAAAATATCGTGTTCCTTA 3112  
QY 61 TATAGATGGAAGCAAGCGAGCAAAAAAGGGAATTTACCTCCGATAGAGGTATACAA 120  
Db 3113 TATAGATGGAAGCAAGCGAGCAAAAAAGGGAATTTACCTCCGATAGAGGTATACAA 3172  
QY 121 GCGTGAAGGATCAATGCTGAGCAAAATGTCATCAAGATTAACAGCAAGCATATGTCCAC 180  
Db 3173 GCGTGAAGGATCAATGCTGAGCAAAATGTCATCAAGATTAACAGCAAGCATATGTCCAC 3232  
QY 181 TTTCAATGGGACACATATATTTATTAATGTTAGAGTTCTTATGACGCTATCATCAG 240  
Db 3233 TTTCAATGGGACACATATATTTATTAATGTTAGAGTTCTTATGACGCTATCATCAG 3292

QY 241 TGAAGATTTACTCATGAAGATCCAACTATATAGCTTAAAGATGAGGATATGTTAATGA 300  
Db 3293 TGAAGATTTACTCATGAAGATCCAACTATATAGCTTAAAGATGAGGATATGTTAATGA 3352  
QY 301 GGTCAGAGGTGAGATATGTTATCAAGGTAGATGAAAAATACTATGTTTAAAGATGC 360  
Db 3353 GGTCAGAGGTGAGATATGTTATCAAGGTAGATGAAAAATACTATGTTTAAAGATGC 3412  
QY 361 TGCCACACGGGATTAAGTCGCTACAAAAAGGAATCAATGCACAAAAACAAGCATAG 420  
Db 3413 TGCCACACGGGATTAAGTCGCTACAAAAAGGAATCAATGCACAAAAACAAGCATAG 3472  
QY 421 TCAACATCGTGAAGGTGGAAGTCCAGAAACGATGCTGCTTCCTTGGACGTTGCA 480  
Db 3473 TCAACATCGTGAAGGTGGAAGTCCAGAAACGATGCTGCTTCCTTGGACGTTGCA 3552  
QY 481 AGGAGCTATATCTACAGATGATGTTATATCTTAAATGCTTCGATATCAATAGAGATAC 540  
Db 3553 AGGAGCTATATCTACAGATGATGTTATATCTTAAATGCTTCGATATCAATAGAGATAC 3592  
QY 541 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACATTTACATTCCTTAAGATGAGTT 600  
Db 3593 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACATTTACATTCCTTAAGATGAGTT 3652  
QY 601 ATCAGTAGGAGATGCTGCTGTCAGAGGCTTCCTATCTGCTGAGAGAAATCTGTCAA 660  
Db 3653 ATCAGTAGGAGATGCTGCTGTCAGAGGCTTCCTATCTGCTGAGAGAAATCTGTCAA 3712  
QY 661 TTCAAGAACCTATTCGCGAGCAAAATATACGATTAACACTTCAAGAAACATGCGTTC 720  
Db 3713 TTCAAGAACCTATTCGCGAGCAAAATATACGATTAACACTTCAAGAAACATGCGTTC 3772  
QY 721 TGTAGCAATTCAGGAATCAAAATACTAAACAGCAACAGCAACACTAACAGTCA 780  
Db 3773 TGTAGCAATTCAGGAATCAAAATACTAAACAGCAACAGCAACACTAACAGTCA 3832  
QY 781 AGCAAGTCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840  
Db 3833 AGCAAGTCAAGATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3892  
QY 841 TCAAGCAATGTAGAAATCTGATGCTGCTTGTGTGATCCAGCAACAAATCAACATGCAAC 900  
Db 3893 TCAAGCAATGTAGAAATCTGATGCTGCTTGTGTGATCCAGCAACAAATCAACATGCAAC 3952  
QY 901 AGCTAGAGGTGTCAGATGCCACAGGAGATCATTTACACTTCACTCTCAAT 960  
Db 3953 AGCTAGAGGTGTCAGATGCCACAGGAGATCATTTACACTTCACTCTCAAT 4012  
QY 961 GTCGATTTGGAAGAAAGCAATGCTGATTTATTCCTTCCTGATGCTTCAACCAATG 1020  
Db 4013 GTCGATTTGGAAGAAAGCAATGCTGATTTATTCCTTCCTGATGCTTCAACCAATG 4072  
QY 1021 GGTACGATTTCAAGGCGAGACCAACCAAGCTCCAGCAACGACCTCCGAGTATGTCAGG 1080  
Db 4073 GGTACGATTTCAAGGCGAGACCAACCAAGCTCCAGCAACGACCTCCGAGTATGTCAGG 4132  
QY 1081 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTGTGTAGTACGTGT 1140  
Db 4133 CCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTGTGTAGTACGTGT 4192  
QY 1141 ACGAAAAAGTGGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200  
Db 4193 ACGAAAAAGTGGGGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4252  
QY 1201 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 1260  
Db 4253 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 4312  
QY 1261 TGTTTTCAACACTTTAACTGCTTAAAAAAGAAATGTTGCTCTGCTGAGCAAGATTTTA 1320  
Db 4313 TGTTTTCAACACTTTAACTGCTTAAAAAAGAAATGTTGCTCTGCTGAGCAAGATTTTA 4372  
QY 1321 TGATTAAGCATATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTGMAATTAAGGCTG 1380

|||||  
Db 4373 TGATAAACCATATATATCTGTTAAAGAGGCTCATAAAGCCTTGTTGAAAAAAGGGTCG 4432  
OY 1381 TAAATCTGATTTCCAGAGCTTAGACAAATTTATAGAACGTTGAATGATGATGACTAA 1440  
Db 4433 TAATTTCTATTTCCAGAGCTTAGACAAATTTATAGAACGTTGAATGATGATGACTAA 4492  
OY 1441 TAAAGAAAAATTTGATGATGATTTATTTGATGATGATGATGATGATGATGATGATG 1500  
Db 4493 TAAAGAAAAATTTGATGATGATTTATTTGATGATGATGATGATGATGATGATGATG 4552  
OY 1501 ACTTGCGCAACCAATTTCTCAATTTGATGATGATGATGATGATGATGATGATGATG 1560  
Db 4553 ACTTGCGCAACCAATTTCTCAATTTGATGATGATGATGATGATGATGATGATGATG 4612  
OY 1561 AGCTGATAGATATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1620  
Db 4613 AGCTGATAGATATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 4672  
OY 1621 TGAAGAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680  
Db 4673 TGAAGAGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4732  
OY 1681 CCTTTCTGATAGGAAAAAGTTGACAGCTCAAGCTATATCTAAAGAAAAAGTATCTACC 1740  
Db 4733 CCTTTCTGATAGGAAAAAGTTGACAGCTCAAGCTATATCTAAAGAAAAAGTATCTACC 4792  
OY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATTCACAGTGAATGATGATGATGATGATG 1800  
Db 4793 TCCATCTCCAGACGAGATGTTAAAGCAATTCACAGTGAATGATGATGATGATGATG 4852  
OY 1801 CAATCGTGTGAAAGGGGAAAAAGCAATTCACAGTGTGCTGCTCATATATGTTGACCA 1860  
Db 4853 CAATCGTGTGAAAGGGGAAAAAGCAATTCACAGTGTGCTGCTCATATATGTTGACCA 4912  
OY 1861 TACAGTTGAGTTAAAAAGCGTAATTTGATGATGATGATGATGATGATGATGATGATG 1920  
Db 4913 TACAGTTGAGTTAAAAAGCGTAATTTGATGATGATGATGATGATGATGATGATGATG 4972  
OY 1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980  
Db 4973 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5032  
OY 1981 TTTGTTTGCAGAGATTAGTACTAGTGAACACCCCTGACGACGATGATGATGATGATG 2040  
Db 5033 TTTGTTTGCAGAGATTAGTACTAGTGAACACCCCTGACGACGATGATGATGATGATG 5092  
OY 2041 TGGATGGGGCAATGCCAGTGAAGCATGTTTGAAGCAAGAAAGCAAGATGATGATGATG 2100  
Db 5093 TGGATGGGGCAATGCCAGTGAAGCATGTTTGAAGCAAGAAAGCAAGATGATGATGATG 5152  
OY 2101 TAAAGAACTTCAAGAGGAGTGAAGAGCCAGTGAAGAAACACCTGCTGAGCAAGATGCC 2160  
Db 5153 TAAAGAACTTCAAGAGGAGTGAAGAGCCAGTGAAGAAACACCTGCTGAGCAAGATGCC 5212  
OY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTTCGCAA 2220  
Db 5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAAGTTTGTTCGCAA 5272  
OY 2221 AGTAACGAGATTTCTAGTCTGAAGAGCAATGCAAGAGAACTTAGTGTTCGCAATTA 2280  
Db 5273 AGTAACGAGATTTCTAGTCTGAAGAGCAATGCAAGAGAACTTAGTGTTCGCAATTA 5332  
OY 2281 TTTGACCTTCGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 2340  
Db 5333 TTTGACCTTCGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 5392  
OY 2341 GTTGTAAAAAGAGATATCTTCATCTGTAAGTAAAGAAAAATTAAC 2389  
Db 5393 GTTGTAAAAAGAGATATCTTCATCTGTAAGTAAAGAAAAATTAAC 5441

RESULT 3

US-08-961-083-65  
; Sequence 65, Application US/08961083  
; Patent No. 6159469  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; COMPUTER: HP Vectra 486/33  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961, 083  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 65:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2290 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-961-083-65  
Query Match 41.38; Score 987.6; DB 3; Length 2290;  
Best Local Similarity 67.7%; Pred. No. 9.6e-262;  
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;  
OY 1 TTTCTAGAGTTGGAGCTGATCAAGCTAGAGCGTTAAGGAAAA--TATGCTGTTTC 57  
Db 4 TTCTATGAACTTGCTGCTCCACCAAGCTGTCAGGTTTAAAGAAAGTCTATATGAGTTTC 63  
OY 58 CTATATAGATGAAAAACAAGCAGCAAAAAAAGGAGAAATTTGACTCTGATGAGTTAG 117  
Db 64 TTATATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 123  
OY 118 CAAGCGTAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGAACCAAGCTATGT 177  
Db 124 TAAAGAGGAGGAGATCAACCGCAAAATNGTATNATCAAGATTACGATCAAGGTTATGT 183  
OY 178 CACTTCATGAGGAGACACATATCATTTATTAATGTAAGGTTCCATTAGCGCTATCAT 237  
Db 184 GACCTTCATGAGACCATTTATCTATTAATGAGCAAGGTTCTTATGATGAGCATCAT 243  
OY 238 CAGTGAAGAAATTAATCTGAAAGATCCAACTATTAAGCTAAAGATGATGATGATGATG 297  
Db 244 CAGTGAAGAGCTCTCTCATGAAAGATCCGAATTTATCACTGTAAGATTCAGACATTTGCAA 303  
OY 298 TGAAGTCAAGGGTGGATGTTTATCAAGGTAAGTGAAGAAATCTATGTTTAACTTAAGA 357  
Db 304 TGAATCAAGGGTGGATGTTTATCAAGGTAAGTGAAGAAATCTATGTTTAACTTAAGA 363  
OY 358 TGTGTCACGAGGATAGCGTCCGTACAAAAAGGAATCAATGAGAAAAACAAGACA 417  
Db 364 TGCAGCTCATCGGATATATTTGCGACAAAAAGAGATTAAAGTCAAGAAAGCAAGCA 423

OY	418	TAGCAGCATCAGTCGAAAGTGGAGTCCAGCAAGAAACGATGGTGGTGGCTTGGACAGCTTC	477
Db	424	CAGTCATATATCAT-----AATCCAAAGACAGATATATGCTGTGGCTGCACGCCAAGCC	474
OY	478	GCAAGACGCTATACTACAGATGATGGTTATATCTTTAATGCTTCGTGATATCATAGAGA	537
Db	475	CCAAGACGTTATATCAACGCGATGATGGTATATCTTCAATGCATCTGATATCATTTGAGA	534
OY	538	TACTGTGATGCTTATATCTGTTCTCTCATGAGATCATTTACATTTACATTTCTTACAGATGA	597
Db	535	CACGGGTGATGCTTATATCTGTTCTCTCACGGCGCACCTTACATTTACATTTCTTAAATGA	594
OY	598	GTTATTCAGCTAGGAGTGGCTGCTGCAGAAAGCCTTCTATCTGCTCAGAGAAATCTCTC	657
Db	595	GTTATTCAGCTAGGAGTGTACTGCTCTCAGAAAGCT-----	629
OY	658	AAATTCAGAACCTATTCGCCGACAAAATAGACATACATTCACAAAGCAAACTGGGTACC	717
Db	630	-----ATTGGAATGGAGAGGAGGATTCGTGCTCTCTTCCATATTTCTAGTTATATAGC	681
OY	718	TTTCTGTAGCAATTCACAGAACTACAAATTCCTAACACAGCAACAAACGACACTAACAG	777
Db	682	AAATTCAGCTCAACCAAGATTGTCCAGAGAACCAATCTGACTGTACATCCCACTTATCA	741
OY	778	TCAAGCAGTCGAAGTATATACATGTATATGTCTTTGAAACAGCTCTCAAACTGCTTT	837
Db	742	TCA-----AAATTCAGGGGAAAACTTTCACAGCCCTTTTACCTGTGAATTTGTCTAAACCTTT	798
OY	838	GAGTCACAGCATATGTAGAAATCTGATGGCTGTGCTTTGATTCAGCAGCAAACTCACAGTCG	897
Db	799	ATCAGAAAGCGCATATGTGGAATCTGATGGCTTTATTTTCAGCCAGGCGCAATCTACAAAGTC	858
OY	898	AACAGCTAGAGGTGTTCAGTCCACAGGAGATCTTACACATTCATCCTTACTCTCA	957
Db	859	AACCGCCAGGGGTGTAGCTGTCCCTCATGTGAACCATTCACACTTTATCCCTTATGAGACA	918
OY	958	AATCTCGAATTTGGAAGAAAGAAATGCTGCTGTATTTATCCCTGCTGTATTCGTTACAAACA	1017
Db	919	AAATCTCGAATTTGGAAGAAAGAAATTCCTGTATTTATCCCTTCTGTATTCGTTACAAACA	978
OY	1018	TTGGGTACGAGTTTCAAGGCCACAGAACACAGTCCACAAACGATCCGGAACCTAGTCC	1077
Db	979	TTGGGTACGAGTTTCAAGAACCCAAACACAGTCCACATTCGATCCGGAACTAGTCC	1038
OY	1078	AGGCCGCAACTCTGCACCAATCTTAAATATAGACTCAATTTCTG-----TTT	1125
Db	1039	AAGTCCCAACTCTGCACCAATCTCTCAACAGATCCCAACATTCATTTGATGAGAAATTT	1098
OY	1126	GGTATGTACACTGGTACGAAAATTTGGGAGAGATATGTATTTGAGAAAAGGCATCTC	1185
Db	1099	GGTCAAAAGAAAGCTGTTCGAAAATATAGCCGATGGTATATCTTTTGGAGGAATTTGAGTTCT	1158
OY	1186	TCGTTATGTCTTTGCCGAAGATTTTACCATCTGAACTGTTAAATCTTTGAAGCAACTT	1245
Db	1159	TCGTTATATCTCCACGCCAAGGATTTTTCACACAGAAACGACGACGACATTTGATGCAACT	1218
OY	1246	ATCAAAAACAAGAGATGTTTTCACACACTTTAACTGTAAAAAAGAAAATGTTGTCTCTCG	1305
Db	1219	GGCCAAACAGAGAAAGTTTATCTATAGCTAGAGAGGTAAAGAAACTGACCTCCATCTAG	1278
OY	1306	TGACCAAGAAATTTTATGATATACCATATATATCTGTTAACTGAGGCTCATTAAGCCTGTCT	1365
Db	1279	TGATCGAGAAATTTTACAAATATAGCTTTATACTTACTAGCAGAAATATCCACAAAGATTACT	1338
OY	1366	TGNAATATAGGAGTCGTAAATTCGATTTTCCAGCCTTAGCAAAATTTATATGAAAGCCTTGA	1425
Db	1339	TGATATATAAAGGTGACAAAGTTGATTTTGAAGCTTTGGATTAACCTGTTGGAACGACTCAA	1398
OY	1426	TGATATATCACTAAATATAAAGAAAATTTGTTAGATATATTTATGGATTCCTACAGCAAT	1485
Db	1399	GGATGTCCAAAGTATATTAAGATCAAGTTATGTGAGNATATATCTTGCTTCTTATAGCTCCGAT	1458
OY	1486	TACCATTCACAGAGACTTGGCAAAACCAATTCATAATTTGATATCTGAGAGCAAGT	1545

Db	1439	TCGTCATCCAGAACGTTTAGSAAACCAAATGGCCAAATTACCTACACTCATGTATGAGAT	1518
Qy	1546	TCGTAATGCTCAATTTAGCTGATATAGTATACACGCTCAGATGGTTTACATTTTGTATGACAA	1605
Db	1519	TCAACTAGCCAAAGTTGGCAGGCAAGTACACAAAGAAAGACGGTTATATCTTTGATCTCTCG	1578
Qy	1606	TGATATATATCGATGTAGAGAGATGCATATGTATACCCCTCATATATGGGCCATAGTACTGT	1665
Db	1579	TGATATTAACCGATATGAGGGGGATGGCTATGTATACCTCCATATATGATACCCATCGACCTG	1638
Qy	1666	GATTGGAAACATATGACGCTTCTCGATTAAGSAAAGTTGACGCTCAACCCATATCTAAAGA	1725
Db	1639	GATTAAAAAATATGTTGTGTCTGAAGCTGAGAGAGCGGACCCAGGCTTATGCTAAAGA	1698
Qy	1726	AAAAGTATCCCTACCTCCATCTCCAGACGCGAGATGTTAAAGCAAAATCAACTGAGAGATAG	1785
Db	1639	GAAAGTTTGAACCCCTCTCTCGACAGACATCAGAGATTTCAAGAAATCTAGGCAAAAGG	1758
Qy	1786	TGCAGACGACTTATTCACATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTGCATTC	1845
Db	1759	AGCAGAAACCTATCTACAAACCGCGTGAAGACAGCTAAAGAGTGCCACTGATGCTATGCGC	1818
Qy	1846	ATATATGTTGATGATACAGTTCAGTTCAGGTTAAACGATTAATTTGATTTTCCCTATAGGA	1905
Db	1819	TTACATATCTTAATATACCTGTAGAAAGTCAAAACGGTATGTTTAATCTATACCTCATATGA	1878
Qy	1906	TCATATACCATATATATTTAAATTTGCTGGTTTGATGATCACACATACAAACCTCCAAATGG	1965
Db	1879	CCATTACCATTAACATCAAAATTTGAGTGGTTTGACGAAGGCCCTTATGACGACCTAAAGG	1938
Qy	1966	CTATATCCTTGAAGATTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTTGACGAAG	2025
Db	1939	GTAATCTCTTGAAGATCTTTTGGCGACTGTCAAGTACTATGTGCAACATCCAAACGAAG	1998
Qy	2026	TCCCATCTCTAATGATGATGGGATGGGCAATCCGATGACCATGTGTTAGCGAAAGAACCA	2085
Db	1999	TCCGATTCACATATATGTTTGGTATAGCGTACGACCATGTTCAAGAAACAAAATTTGG	2058
Qy	2086	CAGTGAAGATCCAATATAGAATCTCAAGCGGATGAGAGACCGAGTAGAGAAACACCTGCG	2145
Db	2059	TCAAGCTGATCCATCATCAACGAAACCAAGAGAGAGAAACCTCAGACAGAAAAACC	2118
Qy	2146	TGAGCCAGAGTCCCTCAAGTAGAGA	2171
Db	2119	TGAGGAAGAAACCCCTCGAGAAAGAGA	2144
RESULT 4			
US-08-961-527-243			
: Sequence 243, Application US/08961527			
: Patent No. 6420135			
GENERAL INFORMATION:			
APPLICANT: Charles Kunsch			
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences			
NUMBER OF SEQUENCES: 391			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Human Genome Sciences, Inc.			
STREET: 9410 Key West Avenue			
CITY: Rockville			
STATE: Maryland			
COUNTRY: USA			
ZIP: 20850			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage			
COMPUTER: HP Vectra 486/33			
OPERATING SYSTEM: MSDOS version 6.2			
SOFTWARE: ASCII Text			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/961,527			
FILING DATE:			
CLASSIFICATION: 424			
PRIOR APPLICATION DATA:			

```

; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 243:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-961-527-243
```

```

Query Match      27.4%; Score 653.8; DB 4; Length 2359;
Best Local Similarity 68.3%; Pred. No. 5.9e-170;
Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;
```

```

QY      1 TTCTTAGAGTTGGAGCTGTTATCAAGCTAGACGGTTAAGGAAA---TAATCGTTTTC 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      939 TTCTTAGAGCTTGGAGCTTACCAAGCTGTCAGAGATAAGAAAGCTATATCAGTTGC 998
QY      58 CTATATAGATGAAGAAACAGCGACCAAAAACGAGAAATTTGACTCTCGTGTAGGTTAG 117
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      999 TTATATATAGATGCTATAGAGCTGTGTCAAAAGCGAAAACCTTGACACAGTGAAGTCAG 1058
QY      118 CAACCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1059 TAAGAGGAGGAGGATCAACGCCCAAAATTTGTTATCAAGTTACGATCAAGGTTATGT 1118
QY      178 CACTTCACATGGCGACCACTATCTTATTAATGTAAGTTCTTCTATGACGTATATC 237
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1119 GACCTGCTATGAGACCATTAATCTATTAATGGAAGGCTTCTTAATGATGCCATCAT 1178
QY      238 CAGTGAAGAAATTAATCAATGAAGATCCAACTATTAAGCTAAAGATGAGGTTATGTTAA 297
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1179 CAGTGAAGAGCTCTCATGAAGATCCGAATTAATCACTGTAAGATTCACAGATTCGCA 1238
QY      298 TGAAGTCAAGGGTGTATGTTATCAAGGTAGATGAAAAATCTATGTATTAACCTTAAGGA 357
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1239 TGAATCAAGGGTGTATGTTATCAATTAAGTAAAGGTAATATCTATGTTAACTTAAGGA 1298
QY      358 TGGTGGCCGAGCGGATATCGCTACAAAAGAGAAATCAATGCAAAAACAAAGCA 417
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1299 TGCAGCTCATGCGGATATATTCGGACAAAAGAGATTAAAGTCAAGACGAGGAACG 1358
QY      418 TAGTCAACATCGTGAAGGTGAATCCCAAGAAACGATGTGCTGTGCTTGGACGTTTC 477
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1359 CAGTCATATATCATATAC-----CAAGACGAGATTAATGCTGCTGTCGACGACAGC 1409
QY      478 GCAAGACGCTATATCAATGATGATGTTATATCTTAATGCTTCTGATATCATAGAGA 537
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1410 CCAAGACGCTTATCAACAGGATGATGGGTATATCTTCAATGCAATCTGATTCATTAAGGA 1469
QY      538 TACTGGTGAATGCTTATATCTGTTCTCATGAGATCATTAACATTACTTCTTAAGATGA 597
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1470 CACGGGTGATGCTTATATCTTCTCATGCGGACCATTAACATTACTTCTTAAGATGA 1529
QY      598 GTTATCAGCTAGCAGATGTTGCTGTCGAGAAGCCTTCCTATGCTGGTGAAGAAATCTGTC 657
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1530 GTTATCAGCTAGCAGATGTTGCTGTCGAGAAGCCTTATGGAATGG----- 1574
QY      658 AAATTCAGAACCTATGCGCAGACAAATATACGATTAACCTTCAAGAACAACTGGGTACC 717
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1575 -GAAGCAGGATCTCGCTTCTCAAGTTCTAGTTATATGCAAAATCCAGCTCAACCA 1633
QY      718 TTCTGTAGCAATTCAGGAAGTACTATACACAGACAAACAGCAACTAAGCAG 777
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1634 GATTGTGAGAGACCAATCTACTCT-----CACTCCAACTTA 1673
```

```

QY      778 TCAAGCAAGTCAAAAGTAATGATGATAGTCTTGAAAACGCTCAACAAGTGCCTT 837
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1674 TCAATCAAAATCAAGGGGAAAACATTTCAAGCCTTTTACGATATGATCTTAACCTTT 1733
QY      838 GAGTCAACGACATGTATAGATCTGATGGCCTTGTCTTTGATCCAGACAAATCAACAGTCG 897
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1734 ATCAAGACGCCATGTGAATCTGATGCGCTTATTTGACCCAGCCGAATATCAAAATGCG 1793
QY      898 AACAGCTAAGAGCTGTTCAGTGGCCACACGAGAGATCATTAACATCTCAATCTACTCTCA 957
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1794 AACCCGACAGGTGTACTCTGCTCCATGATTAACCATTAACCTTATTCCTTATGAA 1853
QY      958 AATGCTGAATTTGGAAGAACGAATCGCTGATATATTCCTCTGTTATGTTCAACCA 1017
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1854 AATGCTGAATTTGGAAGAACGAATGCTGATATATTCCTCTGTTATGTTCAACCA 1913
QY      1018 TTGGGTACAGATTTCAAGGCCAGAACCAACCAAGTCCAAACCGACTCCGAACTACTCC 1077
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1914 TTGGGTACAGATTTCAAGGCCAGAACCAACCAAGTCCAAACCGACTCCGAACTACTCC 1973
QY      1078 AGGCCCCCAACCTGCACCAATCTTAATAATAGACTCAAAATCTTCTTGTGTAGTACGT 1137
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      1974 AAGTCCGCAACAGCTCCAAAGCAATTCATTAATGATGAGAA-----ATTGGTCAAGAAAGC 2027
QY      1138 GGTACGAAAAGTTGGGAGAGATATGATTCGAAGAAAAGGCAATCTGCTATGCTT 1197
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2028 TGTTCGAAAAGTATAGCGATGATGTTATGCTTGTGAGAGAAATGAGATTTCTGCTTATATCC 2087
QY      1198 TGGCAAAAGATTTACCATCTGAACCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAGA 1257
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2088 AGCCAAAGATCTTTACACAGAACAGACAGCATTTGATGAGAACTGCGCAAGCAGGA 2147
QY      1258 GAGTGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTGCTATGACCAAGATT 1317
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2148 AAGTTATCTCATTAAGCTAAGAACTAAGAACTGAACTCCATCTAATGATGAGAAATT 2207
QY      1318 TTATGATTAAGCAATATATCTGTTAACTGAGCTCATTAAGCCTTGTGTTAAATAAGGG 1377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2208 TTCAATTAAGGCTTATATCTTACTACTACAAAGATTTCAACCAAGATTTACTATATAAAG 2267
QY      1378 TCGTAATTCGATTTCCAGCCTTATAGCAAAATTTATGAAGCTTGAAGTATGATGAC 1437
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2268 TCGCAAGTGTATTTGAGCCTTGTGATACCTGTTGGAACGACTCAAGATGTCACAG 2327
QY      1438 TAAATTAAGAAAATTTGTTAGATGATTAATG 1469
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      2328 TGATAAAGTCAAGTTAGTGAAGATATTCTTG 2359
```

```

RESULT 5
US-08-961-083-181
; Sequence 181, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
```







QY 598 GTTATCAGCTAGCGAGTGGCTGCTGCAGAGCCT 632  
|||||  
Db 748 GTTATCAGCTAGCGAGTGGCTGCTGCAGAGCCT 782

## RESULT 8

US-08-961-527-258  
; Sequence 258, Application US/08961527  
; Patent No. 6420135

## GENERAL INFORMATION:

; APPLICANT: Charles Kunach  
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences  
; NUMBER OF SEQUENCES: 391  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA

; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
; OPERATING SYSTEM: MSDOS version 6.2  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/961,527  
; FILING DATE:  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brookes, A. Anders  
; REGISTRATION NUMBER: 36,373  
; REFERENCE/DOCKET NUMBER: PB340P1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (301) 309-8504  
; TELEFAX: (301) 309-8512  
; INFORMATION FOR SEQ ID NO: 258:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1684 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

US-08-961-527-258

Query Match 10.2%; Score 243.4; DB 4; Length 1684;  
Best Local Similarity 60.3%; Pred. No. 3.5e-57;  
Matches 493; Conservative 0; Mismatches 236; Indels 88; Gaps 2;

QY 1629 ATGCATATGTAAGCGCTCATATGAGGCGCATAGTACATGATGGAAAAAGATAGCCTTTG 1688  
|||||  
Db 1 ATGCCATATGTAAGCGCTCATATGAGGCGCATAGTACATGATGGAAAAAGATAGTGTCTG 60  
QY 1689 ATAGGAAAAAGTTGCGATGCAAGCCTATACATAAAGAAAAAGGTATCTACCTCATCTC 1748  
|||||  
Db 61 AAGCTGAGAGAGGCGCA-CCAGGCTTATGCTAAAGAGAAAGTTGACCCCTCTTGA 119  
QY 1749 CAGACGCAATGTTAAAGCAATCACTGAGATAGTGCAGCAGCTATTTTACAAATCGTG 1808  
|||||  
Db 120 CAGACCATGAGATTCAGAAATATGAGGCAAAAGAGAGAGCAAGAGCTATCTCAACCGCG 179  
QY 1809 TGAAGGGGAAAAAGCAATTCACCTGCTTCGATTCATATATGTTGAGCATACAGTTG 1868  
|||||  
Db 180 TGAAGCGACCTAAGAGGTGCGCCTTATGCTATGCTTCAATCTCATATATACGTAG 239  
QY 1869 AGGTAAAAAGCGTATTTGATTTCTCATTAAGATCATTTACATATATTAATTTG 1928  
|||||  
Db 240 AAGTCAAAAAAGGTATTTATATACATCTATATGACCATTTACATACATAAATTTG 299  
QY 1929 CTTGGTTGATGATCAACATATACAAAGCTCAAAATGCTATACCTTGGAAGATTTGTTG 1988  
|||||

Db 300 AGTGGTTTACGAGAGCGCTTTTATAGGCACTTAAGGGATATACCTTTGAGGATCTTTTG 359  
QY 1989 CGACGATTAAGTACTAGTGAACACCTTGACGAGACGTCACATTTCTAATGATGATGGG 2048  
|||||  
Db 360 CGACTGTCAAGTACTATGTGCAACATCCAAACGACGCTCCGCTTGAATATGTTTG 419  
QY 2049 GCAATGCCAGTGAACATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAAATAGAACT 2108  
|||||  
Db 420 GTAACCTGACGACATGTTCAAAGAAACAAANAATGGTCAAGCTATACCAATCAAAAG 479  
QY 2109 TCAAGCGATGA----- 2121  
Db 480 AAAAACCAGCGAGAGAAACCTCAGACAGAAAAACCTGAGAAAGAAACCTCGAGAAAG 539  
QY 2122 -----AGAGCCAGTAGAGAAACAC 2141  
Db 540 AGAACCAGCAAGCGAGAAACAGAGTCTCCAAACCAGAGAGACAGAAAGATCAC 599  
QY 2142 CTGCTGAGCGCAAGTCCCTCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAAAG 2201  
|||||  
Db 600 CAGAGAAATCAGAAAGACCTCAGAGTGAAGAGTTGAAGAAACTGAGAGAG 659  
QY 2202 CAGAGTTTCTGCGAAAGTACGAGATTTAGTCTGAAAGCCAAATGCAACGAAACTC 2261  
|||||  
Db 660 CTGAAAGATTACTTGGAAAAATCCAGAGATCCAAATATCAAGTCAATGCGCAAGAGACTC 719  
QY 2262 TAGCTGTTTACGAATTAATTTGATCTCTCAATTTATGATATCAATATCATAGGACG 2321  
|||||  
Db 720 TCACAGAGATTAATAAATTAATTTACTATTTGGCACCGAGCAACAATATCTATTTGCGAG 779  
QY 2322 AAGCAGAAAAATTAATTTGCTGTTGTTAAAGGAAGTAA 2358  
|||||  
Db 780 AAGCTGAAAAACATATTTGCTTTTAAAGAGAGTAA 816

## RESULT 9

US-08-743-637B-34/C  
; Sequence 34, Application US/08743637B  
; Patent No. 5994066

## GENERAL INFORMATION:

; APPLICANT: BERGERON, Michel G.  
; APPLICANT: PICARD, Francois J.  
; APPLICANT: OUELLETTE, Marc  
; APPLICANT: ROY, Paul H.

; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA

; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND  
IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED

; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...  
; NUMBER OF SEQUENCES: 273  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: OUARLES & BRADY

STREET: 411 EAST WISCONSIN AVENUE

CITY: MILWAUKEE

STATE: WISCONSIN

COUNTRY: USA

ZIP: 53202-4497

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/743,637B

FILING DATE: 04-Nov-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/526,840

FILING DATE: 11-SEP-1995

ATTORNEY/AGENT INFORMATION:

NAME: BAKER, Jean C.

REGISTRATION NUMBER: 35,433

REFERENCE/DOCKET NUMBER: 850586, 90012

TELECOMMUNICATION INFORMATION:



REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMM  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: PTZ9pt-F1s  
US-08-232-463-14

Query Match 2.4%; Score 57.4; DB 1; Length 7218;  
Best Local Similarity 3.4%; Pred. No. 1e-05;  
Matches 13; Conservative 224; Mismatches 150; Indels 0; Gaps 0;

QY 1974 TGGAGATTGTTTGGCATTAACTAGTACAGACACCCCTGACGAGCGTCACATT 2033  
| ||||| | | | :  
DB 1450 TAGAACAATTGGTACRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 1391  
QY 2034 CTAATGATGATGGGCAATGCCAGTACAGTGTGTTAGCAGAGAAAGACACAGTGAAG 2093  
:  
DB 1390 RRR 1331  
QY 2094 ATCCAAATAGAACTTCAAGGGATGAAGAGCAGCTAGAGAAACACTGCTGAGCCAG 2153  
:  
DB 1330 RRR 1271  
QY 2154 AAGTCCCTCAAGTAGAGTGAAGAGCCCACTCAAGAGAGAGAGAGTTTGC 2213  
:  
DB 1270 RRR 1211  
QY 2214 TTGGCAAGTAACGATTTAGTCTGAAGCCATGCAAGACAAGAACTTACGTGTTTAC 2273  
:  
DB 1210 RRR 1151  
QY 2274 GAATTAATTGACTCTTCAATTAATGATACAAATAGTTCATGCGACAGAGAAAAAT 2333  
:  
DB 1150 RRR 1091  
QY 2334 TACTGCGTTGTTAAAGAGAGTAATC 2360  
:  
DB 1090 RRR 1064

RESULT 12  
US-09-134-001C-2243  
Sequence 2243, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucelte-Stamm et al  
TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 2243  
LENGTH: 11091  
TYPE: DNA  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-2243

Query Match 2.2%; Score 51.6; DB 4; Length 11091;  
Best Local Similarity 47.3%; Pred. No. 0.0005;

Matches 187; Conservative 0; Mismatches 205; Indels 3; Gaps 1;

QY 1096 AATCTTAAATAGACTCAATCTCTTGGTTACTACGCTGCTACGAAAGTTGGGA 1155  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8481 AAAAGCAAAATGATGTAATCAATCAAACTAATGACCAAGTGAATCTGAGCA 8540  
QY 1156 AGGATATGTTTCGAAGAAAGGCACTCTGCTATGCTTTGCGAAA---GATTTACC 1212  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8541 AATATTTAGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 8600  
QY 1213 ATCTGAAGCTTTAAATATCTGAAAGCAAGTATCAAAAACAGAGAGTGTTCACAC 1272  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8601 AGCTGAATTTGTTAAAGCCCAACAAACAAATGATGAATAGACCAAGAAATTTAGTC 8660  
QY 1273 TTTACTGCTAAAAAGAAATGTTCTCTCTGCTGACCAAGATTTATGATTAACATA 1332  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8661 TACACAGAGGAAAGAAAGACATGCTTTACAACTTTAGATGAACAGCTTAAGAAATTCAT 8720  
QY 1333 TATCTGTTAACTGAGGCTCATTAAGCCTTGTGTTGNAATTAAGGCTGTAATTCGATTT 1392  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8721 TATCTATTAATCAAGCTAATATACATATATAGATGATTAATGCTAAACCTTGCGTT 8780  
QY 1393 CCAAGCCTTAGACAAATTTATTAAGACGCTTGAATGATGATGATGATGATGATGAT 1452  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 8781 GATTAACATTAAGTATACAGACAGATATATAAAGAAAGAAATGATATTAATTA 8840  
QY 1453 GGATGATGATTTATGTCATTCCTACACCAATTA 1487  
DB 8841 AATGATGTTTCAAGATTCAGAAAGCTATTAATTA 8875

RESULT 13  
US-08-676-967-2  
Sequence 2, Application US/08676967  
Patent No. 5747317  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESS: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,967  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: UCB96-055  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415)343-4341  
TELEFAX: (415)343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2277 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-676-967-2

Query Match 1.9%; Score 45.2; DB 1; Length 2277;  
Best Local Similarity 29.7%; Pred. No. 0.013;



Db 851 SNGARGARGAYWSNGAYYTNGARGARWSNGAYWSNATHGAYGAYGGNGARGARYTNGCNC 910  
OY 519 CTCTGATATCATAGAGA 537  
::||:| :  
Db 911 ARWSNGAYACNWSNACNGA 929

Search completed: November 12, 2002, 15:30:55  
Job time : 141 secs



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 12, 2002, 14:47:29 : Search time 74 Seconds  
(without alignments)  
11453.817 Million cell updates/sec

Title: US-08-961-083-55  
Perfect score: 2389  
Sequence: 1 TTCTTACGAGTTGGACTGT.....TAAGTAAGCAAAATAAAC 2389

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 320260 seqs, 177392727 residues

Total number of hits satisfying chosen parameters: 640520

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_MA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCIT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCIT8\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2388	100.0	2389	10	US-09-765-272-55
2	987.6	41.3	2290	10	US-09-765-272-65
3	385.4	16.1	1342	10	US-09-765-272-181
4	109	4.6	841	10	US-09-452-599-34
5	49.8	2.1	1959	10	US-09-864-761-4012
6	44.6	1.9	423	10	US-09-864-761-18355
7	44.6	1.9	487	10	US-09-864-761-1597
8	43.6	1.8	766	10	US-09-864-761-19608
9	43.6	1.8	1944	10	US-09-864-761-19262
10	42.4	1.8	660	10	US-09-864-761-19488
11	42.4	1.8	962	10	US-09-864-761-2772
12	42.4	1.8	4047	10	US-09-815-242-843
13	42.4	1.8	4050	10	US-09-815-242-9039
14	42	1.8	305	10	US-09-864-761-19262
15	42	1.8	496	10	US-09-864-761-2534
16	41.6	1.7	611	10	US-09-864-864-233
17	41.6	1.7	1845	10	US-09-864-864-213
18	41.6	1.7	1860	10	US-09-919-497-50
19	41.4	1.7	5361	9	US-09-742-096-2

20	41.4	1.7	6152	9	US-09-742-096-1	Sequence 1, Appl
21	41.2	1.7	696	10	US-09-922-261-193	Sequence 193, App
22	41.2	1.7	699	10	US-09-922-261-191	Sequence 191, App
23	41.2	1.7	717	10	US-09-922-261-189	Sequence 189, App
24	41.2	1.7	774	10	US-09-922-261-187	Sequence 187, App
25	41.2	1.7	819	10	US-09-922-261-185	Sequence 184, App
26	41.2	1.7	1669	10	US-09-922-261-185	Sequence 184, App
27	40.8	1.7	2000	9	US-09-938-842A-4612	Sequence 4612, App
28	40.6	1.7	275	10	US-09-864-761-20595	Sequence 20595, A
29	40.2	1.7	475	10	US-09-864-761-1619	Sequence 1619, Ap
30	40	1.7	4316	10	US-09-880-107-3713	Sequence 3713, Ap
31	39.6	1.7	1458	10	US-09-815-242-4757	Sequence 4757, Ap
32	39.6	1.7	1599	10	US-09-815-242-4824	Sequence 8824, Ap
33	39.4	1.6	315	10	US-09-864-761-21723	Sequence 21723, A
34	39.2	1.6	462	10	US-09-864-761-1829	Sequence 3829, Ap
35	39.2	1.6	1074	10	US-09-861-451A-49	Sequence 49, Appl
36	39	1.6	477	10	US-09-864-761-5436	Sequence 5436, Ap
37	38.8	1.6	277	10	US-09-294-093B-1831	Sequence 1831, Ap
38	38.8	1.6	505	10	US-09-917-800A-314	Sequence 314, Ap
39	38.8	1.6	510	10	US-09-864-761-18737	Sequence 18737, A
40	38.8	1.6	2150	10	US-09-826-752-13	Sequence 13, Appl
41	38.2	1.6	1002	12	US-10-007-693-119	Sequence 119, App
42	38.2	1.6	2577	12	US-10-007-693-12	Sequence 32, Appl
43	37.8	1.6	184	10	US-09-864-761-20375	Sequence 20375, A
44	37.8	1.6	489	10	US-09-864-761-1607	Sequence 1607, Ap
45	37.8	1.6	583	10	US-09-864-761-20772	Sequence 20772, A

#### ALIGNMENTS

RESULT 1  
US-09-765-272-55  
Sequence 55, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
Applicant: Chai et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765, 272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961, 083  
FILING DATE: <unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 55:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2389 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 55:  
US-09-765-272-55

Query Match	100.0%;	Score 2388;	DB 10;	Length 2389;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2389;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

OY	1	TTTTCACAGTGTGGGACTGTATCTACAGTATACAGGTTAAGGAAATATATCGTGTTCCTA	60
Db	1	TTCTTCACAGTGTGGGACTGTATCTACAGTATACAGGTTAAGGAAATATATCGTGTTCCTA	60
OY	61	TATAGATGGAATACAGCAGCAGCAAAAGAGAGATTTGACTCCGTGATGAGTATAGCA	120
Db	61	TATAGATGGAATACAGCAGCAGCAAAAGAGAGATTTGACTCCGTGATGAGTATAGCA	120
OY	121	GCGTGAAGGAATCAATGCTGAGCAATCGTATCAAGATPACAGACCAAGGCTATGTAC	180
Db	121	GCGTGAAGGAATCAATGCTGAGCAATCGTATCAAGATPACAGACCAAGGCTATGTAC	180
OY	181	TTTCACATGGGACACATATCATTTATTCATAGGTAAAGTTCCTTATGACGCTATCATAG	240
Db	181	TTTCACATGGGACACACATATCATTTATTCATAGGTAAAGTTCCTTATGACGCTATCATAG	240
OY	241	TGAAGAAATTACTCATGAAAGATCCAAACTTAAAGCTPAAAGATGAGAGATTTGTTAAAGA	300
Db	241	TGAAGAAATTACTCATGAAAGATCCAAACTTAAAGCTPAAAGATGAGAGATTTGTTAAAGA	300
OY	301	GGTCAAGGCTGGATATGTATTTATTCAAAGTATGGAATATCTATGTTTACCTTAAAGATGC	360
Db	301	GGTCAAGGCTGGATATGTATTTATTCAAAGTATGGAATATCTATGTTTACCTTAAAGATGC	360
OY	361	TGCCACACGCGATTAACGTCGTCACAAAGAGAAATCAATCGACAAAACAGAGCATATG	420
Db	361	TGCCACACGCGATTAACGTCGTCACAAAGAGAAATCAATCGACAAAACAGAGCATATG	420
OY	421	TCACATCTGTAGAGGTGAATCCACAGAAACGATGTGCTGTTCCTTGGCAGCTTGCCTA	480
Db	421	TCACATCTGTAGAGGTGAATCCACAGAAACGATGTGCTGTTCCTTGGCAGCTTGCCTA	480
OY	481	AGGAGCGATATCTACATATGATATGTTATATCTTATGCTCTGATATCATATGAGATATC	540
Db	481	AGGAGCGATATCTACATATGATATGTTATATCTTATGCTCTGATATCATATGAGATATC	540
OY	541	TGGTGATGCTTATATTCGTTCTTCATGAGATCATCTTACCATTTCCTAAGAAATGAGATT	600
Db	541	TGGTGATGCTTATATTCGTTCTTCATGAGATCATCTTACCATTTCCTAAGAAATGAGATT	600
OY	601	ATTCAGCTATGCGAGATTTGGCTGCTGCACAGACGCTTCTATCTGCTGAGGAAATCTGTCAA	660
Db	601	ATTCAGCTATGCGAGATTTGGCTGCTGCACAGACGCTTCTATCTGCTGAGGAAATCTGTCAA	660
OY	661	TTTCAAGAACCTATGGCGGACAAAATATGCGATTAACCTTAAAGAACAACTGGGTACCTTC	720
Db	661	TTTCAAGAACCTATGGCGGACAAAATATGCGATTAACCTTAAAGAACAACTGGGTACCTTC	720
OY	721	TGTAAAGCAATCCAGAACTACAAATACTAACACAGCAACAACAGCAACACTTACAGTCA	780
Db	721	TGTAAAGCAATCCAGAACTACAAATACTAACACAGCAACAACAGCAACACTTACAGTCA	780
OY	781	AGCAAGTCAAAAGTATATACATTTATATGTCTCTTGAACAGCTCTACAAATCTGCTTTAG	840
Db	781	AGCAAGTCAAAAGTATATACATTTATATGTCTCTTGAACAGCTCTACAAATCTGCTTTAG	840
OY	841	TCAACGACATATTAACATCTATATGAGCTTGTCTTTGATTCAGGACAAATATCAGAGTGCAC	900
Db	841	TCAACGACATATTAACATCTATATGAGCTTGTCTTTGATTCAGGACAAATATCAGAGTGCAC	900
OY	901	AGCTAGAGGTGTTCAGTGCACACGAGATCATTTACCACTTCATCCCTTACTGTCAAT	960
Db	901	AGCTAGAGGTGTTCAGTGCACACGAGATCATTTACCACTTCATCCCTTACTGTCAAT	960
OY	961	GTCCTGAATTGGAGAACGAATGCGCTGTATATTTCCCTTCTGTATTCGTTCAAAACCATTG	1020
Db	961	GTCCTGAATTGGAGAACGAATGCGCTGTATATTTCCCTTCTGTATTCGTTCAAAACCATTG	1020

QY	1021	GGTACACGATTTCAAGGCCAGACACCAGCTCCACACACCGACTCCGGAGACCTGATGCCAG	1080
Db	1021	GGTACACGATTTCAAGGCCAGACACCAGCTCCACACACCGACTCCGGAGACCTGATGCCAG	1080
QY	1081	CCCCGACCTCCACCAATGCTTAAAAATGACCTCAATTTCTCTTGTTGGTTAGTACGCTGGT	1140
Db	1081	CCCCGACCTCCACCAATGCTTAAAAATGACCTCAATTTCTCTTGTTGGTTAGTACGCTGGT	1140
QY	1141	ACGAAAGCTTGGGGAGAGATATGATTCGAGAGAAAAGGGCATCTCTCGTTATATGCTTTGGC	1200
Db	1141	ACGAAAGCTTGGGGAGAGATATGATTCGAGAGAAAAGGGCATCTCTCGTTATATGCTTTGGC	1200
QY	1201	GAAGATTTACCACTCTGAAACCTGTAAAAATCTTTGGAAGCAAGTTATCAAAACAAGAG	1260
Db	1201	GAAGATTTACCACTCTGAAACCTGTAAAAATCTTTGGAAGCAAGTTATCAAAACAAGAG	1260
QY	1261	TGTTTCACACGCTTTAACTCTGTAAAAAGAAAATGTGCTCTCGTGGACCAAGAAATTTTA	1320
Db	1261	TGTTTCACACGCTTTAACTCTGTAAAAAGAAAATGTGCTCTCGTGGACCAAGAAATTTTA	1320
QY	1321	TGATTAAGCATATATCTGTTTAACTGAGGCTCTAAAGGCTTGTTTGNAATTAAGGGTGC	1380
Db	1321	TGATTAAGCATATATCTGTTTAACTGAGGCTCTAAAGGCTTGTTTGNAATTAAGGGTGC	1380
QY	1381	TAATTCCTGATTTCCAGCCTTAGACAAATTTATTAGACGCTTGAATGATGCATTA	1440
Db	1381	TAATTCCTGATTTCCAGCCTTAGACAAATTTATTAGACGCTTGAATGATGCATTA	1440
QY	1441	TAAAGAAAAATTGGTAGATGTTTAAATGGGATTCCTGAGACCAATTTACCATCCAGACG	1500
Db	1441	TAAAGAAAAATTGGTAGATGTTTAAATGGGATTCCTGAGACCAATTTACCATCCAGACG	1500
QY	1501	ACTTGGCAAAACCAATTCCTCAAAATGAGTATATCGAAGAGCAAGTTCGTATTTGCTCAAT	1560
Db	1501	ACTTGGCAAAACCAATTCCTCAAAATGAGTATATCGAAGAGCAAGTTCGTATTTGCTCAAT	1560
QY	1561	AGCTGATATGATATACAACGTCAGATGGTTACATTTTGTATGAACATGATATATACGTGA	1620
Db	1561	AGCTGATATGATATACAACGTCAGATGGTTACATTTTGTATGAACATGATATATACGTGA	1620
QY	1621	TGAAGGAACTCATATGTAAAGGCTCAATATGGGCGCATCTGACTGGATTTGGAAAAAGATG	1680
Db	1621	TGAAGGAACTCATATGTAAAGGCTCAATATGGGCGCATCTGACTGGATTTGGAAAAAGATG	1680
QY	1681	CCTTTCCTGATTAAGGAAAAAGTTGCAGCTCAAGCTATATCAATAAGAAAAAGTATCTTACC	1740
Db	1681	CCTTTCCTGATTAAGGAAAAAGTTGCAGCTCAAGCTATATCAATAAGAAAAAGTATCTTACC	1740
QY	1741	TCCATCTTCAGACGCGAGATGTTAAAGCAATTCACCTGAGATAGTGCAGCAGCTATTTA	1800
Db	1741	TCCATCTTCAGACGCGAGATGTTAAAGCAATTCACCTGAGATAGTGCAGCAGCTATTTA	1800
QY	1801	CAATCGTGTAAGGAGAAAAAGCAATTCACGCTGACCTTCATATATGTGTGGACA	1860
Db	1801	CAATCGTGTAAGGAGAAAAAGCAATTCACGCTGACCTTCATATATGTGTGGACA	1860
QY	1861	TACAGTTAGGTTAAAAACGTAATTGATATTTCTCTTAAGATCAATTACCAATTAAT	1920
Db	1861	TACAGTTAGGTTAAAAACGTAATTGATATTTCTCTCTTAAGATCAATTACCAATTAAT	1920
QY	1921	TAAATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTTATACCTTGGAGA	1980
Db	1921	TAAATTTGCTTGTGTTGATGATCACACATCAAAAGCTCCAAATGGCTTATACCTTGGAGA	1980
QY	1981	TTTGTGTTGCGACGATTTAAGTACTACTAAGAACCCCTGACGAACTCCACATTTCTAATGA	2040
Db	1981	TTTGTGTTGCGACGATTTAAGTACTACTAAGAACCCCTGACGAACTCCACATTTCTAATGA	2040
QY	2041	TGGATGGGGGAATGGCAGTAGAGATGTTTAAAGCAAGAAACCAACATGAAATTCACAA	2100
Db	2041	TGGATGGGGGAATGGCAGTAGAGATGTTTAAAGCAAGAAACCAACATGAAATTCACAA	2100
QY	2101	TAAACATTTCAAAAGCGATGAAAGCCACTAGAGAAACACCTCGTGAGCCAGATGCC	2160

|||||  
Db 2101 TAAGACTTCAAAAGCGATGAGAGAGCCAGTAGAGGAAACACCTGCGAGCCAGCAACTGCC 2160  
QY 2161 TCAGTAGAGACTGAAAAAGTAGAACCCCACTCAAAAGACGAAAGTTTGGCTTGC 2220  
|||||  
Db 2161 TCAGTAGAGAGACTGAAAAAGTAGAACCCCACTCAAAAGACGAAAGTTTGGCTTGC 2220  
QY 2221 AGTAAGGATTTGACTGCTGAAAGCCAAATGCAACAGAACTCTGCTGCTTTCGAAATA 2280  
|||||  
Db 2221 AGTAAGGATTTGACTGCTGAAAGCCAAATGCAACAGAACTCTGCTGCTTTCGAAATA 2280  
QY 2281 TTTGACTCTTCAAAATTTATGATGATTAACAATAGTATCATGCGAAGACGAAATTTACTTGC 2340  
|||||  
Db 2281 TTTGACTCTTCAAAATTTATGATGATTAACAATAGTATCATGCGAAGACGAAATTTACTTGC 2340  
QY 2341 GTTGTAAAAAGAGATATCTTCTCATCTGTATAGTAAGAAAAAATTAAC 2389  
|||||  
Db 2341 GTTGTAAAAAGAGATATCTTCTCATCTGTATAGTAAGAAAAAATTAAC 2389  
RESULT 2  
US-09-765-272-65  
Sequence 65, Application US/09765272  
Patent No. US2002061545A1  
GENERAL INFORMATION:  
Applicant: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 65:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2290 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 65:  
US-09-765-272-65  
Query Match 41.3% Score 987.6; DB 10; Length 2290;  
Best Local Similarity 67.7% Pred. No. 2.4e-240;  
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;

Db 64 TTATATAGATGGTATGATCAGGCTGTGCAAAAAGCAGAAAACTTACACACATGATGAC 123  
QY 118 CAGCGTGAAGGAATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGACCAGGCTATGT 177  
|||||  
Db 124 TAAGAGGAGGAGGATCAACGCCGCAAAATNGTATCAAGATTAACGATCAAGATTATGT 183  
QY 178 CACTTCAATGGGAGCACCATCATATTATTAACAATGGTAAGTTCCTTATGACCTATCAT 237  
|||||  
Db 184 GACCTCTATGAGAGACCATTAATCATTAATAAGTTCGCAAGGTTCTTATGATGCCATCAT 243  
QY 238 CAGTGAAGATTTACTCATGAAGATCCAAACTATTAAGCTAAAAGATGAGAGATTTGTTAA 297  
|||||  
Db 244 CAGTGAAGAGCTCTCATGAAGATCCGAAATTTTCAGTTGAAGATTTGACATTTGTCAA 303  
QY 298 TGAAGTCAAGGGTGTATGTTATATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGA 357  
|||||  
Db 304 TGAATCAAGGGGTGTATGTTATATCAAGGTAGATGAAAAATCTATGTTTACCTTAAGA 363  
QY 358 TGGTCCCAAGGGGATTAAGCTCCGTACAAAAGAGAAATCAATGCAAAAAGCAAGCA 417  
|||||  
Db 364 TGCAGCTCATGCGGATTAATTTGCAAAAAGAGATTAAACGTGAGAAAGCAAGCAAG 423  
QY 418 TAGTCAACATCGTGAAGGTGAACTCCAGAAACGATGGTGTGCTTGGCAGCTTC 477  
|||||  
Db 424 CAGTCAATATCAT-----AATCAAGAGCAGATTAATGCTGTGCTCCAGCCAGAGC 474  
QY 478 GCAAGGACCTTACTATCAAGATGATGTTATATCTTTAATGCTTCTGATATCATAGGA 537  
|||||  
Db 475 CCAAGGACCTTATATCAACGATGATGTTATATCTTCAATGATCATGATATCATGAGGA 534  
QY 538 TACTGGTATGCTTATATGTTTCCATGATGAGATGATTCATTAATTCCTTAAGATGA 597  
|||||  
Db 535 CAGGGTATGCTTATATGTTTCTCAGCGCAGCATTAATTAATTCCTTAAGATGA 594  
QY 598 GTTATCAGCTAGAGATGTTGCTGCTCAGAGGCTTCTATCTGTGCGAGAAATCTGTC 657  
|||||  
Db 595 GTTATCAGCTAGAGATGTTGCTGCTCAGAGGCTTCTATCTGTGCGAGAAATCTGTC 629  
QY 658 AATTCAGAACCTTATGCGCCAGCAAAATAGCATTAACCTTCAAGAACTGGGTACC 717  
|||||  
Db 630 -----ATTGAAATGGAAGACAGGATCTGCTCTTCAAGTTCTATGTTAATGC 681  
QY 718 TTCTGTAAACCAATCAAGAACTAACAATTAACAACAACAACAACAACAACAACAACA 777  
|||||  
Db 682 AATTCAGCTCAACCAAGATTTGTCAGAGAACCAATTCGACTGTACTTCAACTATTA 741  
QY 778 TCAGCAAGCTCAAAAGTAAATGATGATGATGCTCTTGAACAGCTCTACAACTGCTTC 837  
|||||  
Db 742 TCA---AATCAAGGGAAGAAACATTTCAAGCCCTTTAGCTGAATTTGATGCTTAACCTT 798  
QY 838 GAGTCAAGCAATGTAGAAATCTGATGCTTGTCTTTGATCCAGCAACAATCAAGATCG 897  
|||||  
Db 799 ATCAGAACCCATGTGGAATCTGATGCTTATTTTCGACCAAGCAATATCAAGATCG 858  
QY 898 AACAGCTAGAGGTGGAGTGGCCAGGAGGAGATTAACCACTTATCCCTTACTCTCA 957  
|||||  
Db 859 AACCCGACAGAGGTGATGCTCTCATGTGTAACCACTTATCCCTTATGTAACCA 918  
QY 958 AATGCTGAATGGAAGAAAGCAATGCTGATTAATTCCTTCTGTTATCTGTTCAACCA 1017  
|||||  
Db 919 AATGCTGAATGGAAGAAAGCAATGCTGATTAATTCCTTCTGTTATCTGTTCAACCA 978  
QY 1018 TTGGGTACCAAGTTTCAAGGCGCAAAACCAACCAAGTCCACACGACTCGGAACTAGTCC 1077  
|||||  
Db 979 TTGGGTACCAAGTTTCAAGGCGCAAAACCAACCAAGTCCACACGACTCGGAACTAGTCC 1038  
QY 1078 AGGCCGCAAGCTTGACCAAAATCTTAAATAGATCAATCTTC-----TTT 1125  
|||||  
Db 1039 AAGTCCGCAAGCTTGACCAAAATCTTCAACCAAGTCCACCAATGATGATGAGAAATTT 1098  
QY 1126 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1185  
|||||  
Db 1099 GGTCAAGAGAGCTGTTCAAGAAAGTAGGCGATGTTATGCTTTGAGGAGCAATGAGCTTC 1158



Db	436	TAATGTTCTGCTAGCAAGAGTCTCAGGACGAPATATACGAACAATGATGGTTATGTTCTTAA	495
OY	517	TGCTTCTATATCATAGAGGACTACTGGTGAATGCTTATATTCGTTCTCATGAGATCANTA	576
Db	496	TCCAGCTCATATTTATCGAAGATAGGGGTAAGCTTATATCTTCCTCATGAGAGGACACTA	555
OY	577	CCATTACATTCCTAAGATGAGTATATCAGTACGAGAGTGGCTGGCTGCAGAGACCTTCCT	636
Db	556	TCACTACATTTCCCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTAAAGCACATCT	615
OY	637	ATCTGCTGAGGAAATCTGTCAAAATTTCAAGAACCATTATGCCGACAAATAATGCGATTAACAC	696
Db	616	GGCTGGAATAAATATGCAACCGAGTCATTTAGCTATTCTTCAACAGCTGAGACAA---	672
OY	697	TTCAAGAACCAACTGGTACTCTTCTGTAGCAATTCAGGAACACTACAATTAACACAAAG	756
Db	673	-----TAAACGCGCATTCGT	687
OY	757	CAACACAGCAACACTAATCAGTCAAGCAAGTCAAGTAAATGATGATTAATGCTTTGAA	816
Db	688	AGCAAAAGATCACTACTACAGCAAGCCAGCAATTAATATCTAAAAATCTCCAGAGCTTTTGA	747
OY	817	ACAGCTTCAACAACTGCTTTGATGATCAACGACATGTAGATGATGATGGCTTGTTTGA	876
Db	748	GGAATCTATATATTCACCTAGCGCCCAACGTTACAGTAAATACAGATGGCTGTTTGA	807
OY	877	TTCACACAAATTCACACTGACAGCGATGAGGTGTCACATGCCACAGCGAGATCATTA	936
Db	808	CCCTGCTAAGATTTATCACTGCTGACCCCAATGAGATGCGATTCGCAATGGGACCATTA	867
OY	937	CCACTTCATCCCTTACTCTCAAAATGTCTGAATTTGGAAAGAACGAATGCTGATTAATTC	996
Db	868	CCACTTATTCCTTACAGCAAGCTTTCTGCCCTAGAGAAAAAGATTGCCAGATGTGCC	927
OY	997	CCTTCCT 1003	
Db	928	TATCAGT 934	

Query Match	4.68;	Score 109;	DB 10;	Length 841;
Best Local Similarity	65.48;	Pred. No. 4.3e-18;		
Matches 176;	Conservative 0;	Fragments 90;	Indels 3;	Gaps 1.
0y	2090	GAAGATCCCAATTAAGAACTTCAAAACCGGATGAAGAGCCAGTGAAGAGAAACACCTGCTGAG	2149	

Db	633	GAGTCTCCAAAACCCAGACAGAGAAACCGAAGAAAGATTCACACGAAAGAAATCACCACAGAGAA	574
Oy	2150	CCAGAAATGCTCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAACAAAGTT	2209
Db	573	TCACAAAGAACTCTCGATCGAGACTGAAAGGTTTAGAA--AACTGAGCAAGGCTCAAAAT	517
Oy	2210	TTGCTTGAGAAAGTAGACGGAATTCTAGCTTAAGCCCAATGCACAGAAACTCTAGCTGGT	2268
Db	516	TTACTTGGAATAATCCAGATCCCATTTATCAATCCATCCCAAGSAGACTCTCACAGAA	457
Oy	2270	TTACGAATAATTTGACTCTTCAAATTAGGATACAAATAGTATCATGCGCAGACAGAA	2329
Db	456	TTAAAAAATAATTACTATTGTCGACCAGACACAAATACATATTATGCGCAGAACTGAA	397
Oy	2330	AAATTACTGCGTGTAAAGAGATTA	2358
Db	396	AACTATTGGCTTTATTAAAGGAGACTTA	368

```

RESULT 5
US-09-864-761-4012
; Sequence 4012, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 4012
;
; LENGTH: 1959

```

```

: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL022334.1
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.9
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.1
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.3
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.4
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.3
US-09-864-761-4012

Query March 2.1%, Score 49.8, DB 10, Length 1959;
Best Local Similarity 50.5%, Fred. No. 0.0063;
Matches 148; Conservative 0; Mismatches 142; Indels 3; Gaps 1;
```

	Query Match	Best Local Similarity	2.18	Score 49.8	DB 10:	Length 1959;
	Matches 146;	Conservative	0;	Mismatches 142;	Indels 3;	Gaps
OY	279	AAGATGACGATAVTTGTTAATGAGTCAAGGCGTGATATTGTATCAAAGTATGAGAAAT	338	                               		
Db	478	ATCATGATGATGGTGGTAGTGATGGTGAAGGTGGTGATGATTAATGGATGATGATGATC	537	                               		
OY	339	ACTATGTTACTCTTAAGSAGCTCCACCACCGGTAACGCCCTACAAAAGAGAATCA	398	                               		
Db	538	ATGGTGATGATGGGAATGATGTCATGTCATGTCATGCACAGATGATGATGATGTC	597	                               		
OY	399	ATCGACAAAAACAAGACCATAGTCAACATCGTGAAGGTGGAACGCCAAGAACAGTGTG	458	                               		
Db	598	ATGATGGGGGATGATGACAGACGTGATGTCGTGATGTCGTGATGATGATGATGATA	657	                               		
OY	459	CTGTGGCTTTGGCCGCTTCGCAAGACGCATPACTACAGATGATGGTATATCTTTAAAG	518	                               		
Db	658	GTCATGATGGGGAAGATG---ATCACACGTGATGGTGTGATGATGACAGATAGCGATGATG	714	                               		
OY	519	CTTCGTATATCATAGACGATACGTGTCATCTTATATCGTTCCATGAGAT	571	                               		
Db	715	GGAGATGTGACAGATGATGTCATGATGTCATGATGATGATGATGATGATGATGATGAT	767	                               		

RESULT 6  
US-09-864-761-18355  
Sequence 18355, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE REFERENCE: Aemica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIORITY APPLICATION NUMBER: US 60/180,312  
PRIORITY FILING DATE: 2000-02-04  
PRIORITY APPLICATION NUMBER: US 60/207,456  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: US 09/632,366  
PRIORITY FILING DATE: 2000-08-03  
PRIORITY APPLICATION NUMBER: GB 24263.6  
PRIORITY FILING DATE: 2000-10-04  
PRIORITY APPLICATION NUMBER: US 60/236,359  
PRIORITY FILING DATE: 2000-09-27  
PRIORITY APPLICATION NUMBER: PCT/US01/00666  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00667  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00664  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00669  
PRIORITY FILING DATE: 2001-01-30

```

1  PRIOR APPLICATION NUMBER: PCT/US01/00665
2  PRIOR FILING DATE: 2001-01-30
3  PRIOR APPLICATION NUMBER: PCT/US01/00668
4  PRIOR FILING DATE: 2001-01-30
5  PRIOR APPLICATION NUMBER: PCT/US01/00663
6  PRIOR FILING DATE: 2001-01-30
7  PRIOR APPLICATION NUMBER: PCT/US01/00662
8  PRIOR FILING DATE: 2001-01-30
9  PRIOR APPLICATION NUMBER: PCT/US01/00661
10 PRIOR FILING DATE: 2001-01-30
11 PRIOR APPLICATION NUMBER: PCT/US01/00670
12 PRIOR FILING DATE: 2001-01-30
13 PRIOR APPLICATION NUMBER: US 60/234,687
14 PRIOR FILING DATE: 2000-09-21
15 PRIOR APPLICATION NUMBER: US 09/608,408
16 PRIOR FILING DATE: 2000-06-30
17 PRIOR APPLICATION NUMBER: US 09/774,203
18 PRIOR FILING DATE: 2001-01-29
19 NUMBER OF SEQ ID NOS: 49117
20 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
21 SEQ ID NO 18355
22 LENGTH: 423
23 TYPE: DNA
24 ORGANISM: Homo sapiens
25 FEATURE:
26 OTHER INFORMATION: MAP TO AC010133.1
27 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
28 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
29 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
30 OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 4.2
31 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
32 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
33 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
34 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
35 OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
36 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
37 US-09-864-761-18355

```

	Query Match	1.98	Score 44.6	DB 10	Length 423	
	Best Local Similarity	51.88	Pred. No. 0.062	Mismatches 94	Indels 0	Gaps -
	Matches 101	Conservative	0			
QY	2012	CACCTGACCAACGTCACATTCCTAATGATGATGGCGGCAMTGCAGTCAGCTGTGTTA	2071	11	11	11
Db	7	CTCTGACACAAATGGCGAGTCTTCTGAAAAGAGAGGCGAGCAAAAAGAAAGATGTGAG	66	11	11	11
QY	2072	GGCAAGAAAGCCACAGTGAAGTCCAAATAGAACTTAAAGCGGATGAGAGCCAGTA	2131	11	11	11
Db	67	GAGGAGGAAAGAAAGAGGAGGAGGAGGAGGAGAGCAAGAAAGAAAGAACMACMAA	126	11	11	11
QY	2132	GAGGAAACACCTGCTGAGCCAGCAAGTCCCTCAAGTGAAGTGAAGTGAAGACCCCA	2191	11	11	11
Db	127	GAAACAAGAAAGAGAACAAAGAAAGAAAGAAACAAACAAACAAAGAAAGAACACAGA	186	11	11	11
QY	2192	CTCAAGAGAGCAGAA	2206	11	11	11
Db	187	CACCAAGAAAGAGAA	201	11	11	11

```

RESULT 7
US-09-864-761-1597
? Sequence 1597, Application US/09864761
? Patent No. US20020048763A1
? GENERAL INFORMATION:
? APPLICANT: Penn, Sharron G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? APPLICANT: Chen, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
? FILE REFERENCE: Aecomica-X-1
? CURRENT APPLICATION NUMBER: US/09/864,761
? CURRENT FILING DATE: 2001-05-23

```



```

PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 1597
LENGTH: 487
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010133.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
US-09-864-761-1597

Query Match      1.9%; Score 44.6; DB 10; Length 487;
Best Local Similarity 51.8%; Pred. No. 0.066; Indels 0; Gaps 0;
Matches 101; Conservative 0; Mismatches 94;

QY 2012 CACCTGAGCAAGCTCCATCTATGATGGGCAATGCCAGTACGATGTGA 2071
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 166 CTCTGAGCAATGGCAGATTTCTGAAGAGGAGAGAGAGAGAGAGAGAG 225
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 2072 GCGAAGAAGACCAAGTGAAGTCCAAATTAAGAACTTCAAGCGGATGAAGCCAGTA 2131
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 226 GAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 285
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 2132 GAGGAACACCTGCTGAGCCAGAACTCCCTCAAGTAGACAGCTGAAAAAGTAAAGCCCAA 2191
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
DB 286 GAACAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 345
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
QY 2192 CTCGAAGAAGCAGAA 2206
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```

DB 346 CAACAAGAGAGAGAA 360
| 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
RESULT 8
US-09-864-761-19608/c
Sequence 19608, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecmics-X-1
CURRENT FILING DATE: US/09/864,761
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 19608
LENGTH: 766
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL008720.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
OTHER INFORMATION: SWISSPROT HIT: P17164, EVALU 4.10e+00
OTHER INFORMATION: EST_HUMAN HIT: AW844901.1, EVALU 2.90e+00
US-09-864-761-19608

Query Match      1.8%; Score 43.6; DB 10; Length 766;
```

Best Local Similarity 46.4%; Pred. No. 0.15;  
Matches 180; Conservative 0; Mismatches 204; Indels 4; Gaps 1;

```
OY 186 ATGGCGACCCCTTCATATATACATGTTAGTTCCTTATGACGCTATCATCAGTGAAG 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 429 ATGAAAATGATGATGACGGGATGATGAGATGATATAATGATGATGATGATG 370
OY 246 AATTACTCATGAAGAATCCAACTAAGCTAATAAGATGAGATATGTTAATGAGGTCA 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 369 ATTATGATATGATGATGAGGAGATGAGAGATGAGATGATGATGATGATGATTA 310
OY 306 AGGCTGATATGTTATCAAGCTAGATGGAATAATACATGTTTACCTTAAGAGTGTCCC 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 309 ---TGATGCTGATGAGAGAGATGCGCATGATGCTGATGATGATGATGATGATGCGC 254
OY 366 ACCGGATTAACGTCCTGACAAAGAGAAATCAATGACAAAACAAAGCATATGCAAC 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 253 ATCAAGATGACGATGATGAGAGAGAGGTTGGTGTGATGATGATGATGATGATG 194
OY 426 ATCGTAAGTGGGAACTCCAAAGAACGATGCTGTGCTTGCCCTTGCAAGGAC 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 134
OY 486 GCTATCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 TTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 74
OY 546 ATGCTTATATCGTTCCTCAATGAGATCA 573
DB 73 AAGAGATGCTGATGATGATGATGATGA 46
```

## RESULT 9

```
US-09-864-761-2825/C
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2825
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825
Query Match
Best Local Similarity 1.8%; Score 43.6; DB 10; Length 1944;
Matches 180; Conservative 0; Mismatches 204; Indels 4; Gaps 1;
```

```
OY 186 ATGGCGACCCATATATATACATGTTAGTTCCTTATGACGCTATCATCAGTGAAG 245
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 666 ATGAAAATGATGATGACGGGATGATGAGATGATGATGATGATGATGATGATG 607
OY 246 AATTACTCATGAAGAATCCAACTATAACCTAAGATAGATATGTTAATGAGGTCA 305
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 606 ATTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 547
OY 306 AGGCTGATATGTTATCAAGCTAGATGGAATAATACATGTTTACCTTAAGATGCTGCC 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 546 ---TGATGCTGATGAGAGAGAGATGCGCATGATGCTGATGATGATGATGATG 491
OY 366 ACCGGATTAACGTCCTGACAAAGAGAAATCAATGACAAAACAAAGCATATGCAAC 425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 490 ATCAAGATGACGATGATGAGAGAGAGGTTGGTGTGATGATGATGATGATGATG 431
OY 426 ATCGTAAGTGGGAACTCCAAAGAACGATGCTGTGCTTGCCCTTGCAAGGAC 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 430 GTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 371
OY 486 GCTATCTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 TTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 311
OY 546 ATGCTTATATCGTTCCTCATGAGATCA 573
DB 310 AAGAGATGCTGATGATGATGATGATGA 283
```

## RESULT 10

```
US-09-864-761-19488
; Sequence 19488, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
```

```

? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: US 60/207,456
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00667
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 19488
? LENGTH: 660
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC008125.9
? OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
? OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 3.8
? OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3
? OTHER INFORMATION: NT HIT: AF044255.1, EVALUATE 1.10e-01
US-09-864-761-19488
```

```

Query Match 1.8%; Score 42.4; DB 10; Length 660;
Best Local Similarity 48.7%; Pred. No. 0.28;
Matches 146; Conservative 0; Mismatches 151; Indels 3; Gaps 1;
```

```

QY 278 AAGATGAGGATATGTTAATGAGTCAAGGTGATGTTATCAAGTAGATGGAAGA 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 108 AATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 338 TACTATGTTTACCTTAAGATGCTGCCACGGGATACGTCCTCAAAAGGAATC 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 168 GGTATTAGGGTGGATGATCTATATGGTGTGATGATGATGATGATGATGATGCT 227
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 AATGCACAAAACAGAGAGATGATCAACATCGTGAAG--TGGAACCTCAGAAACGAT 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 228 GATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 455 GGTCCTGCTTCCCTTGCGACGTTGCGCAAGAGCCCTATACTACAGATGATGTTATCTTT 514
```

```

DB 288 AGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 515 AATGCTTCGATATCATATAGGATGATGATGATGATGATGATGATGATGATGAT 574
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 348 GAGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 407
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-09-864-761-2772
? Sequence 2772: Application US/09864761
? Patent No. US20020048763A1
? GENERAL INFORMATION:
? APPLICANT: Penn, Sharon G.
? APPLICANT: Rank, David R.
? APPLICANT: Hanzel, David K.
? APPLICANT: Chen, Wensheng
? TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
? FILE REFERENCE: Aecmca-x-1
? CURRENT FILING DATE: 2001-05-23
? PRIOR APPLICATION NUMBER: US 60/180,312
? PRIOR FILING DATE: 2000-02-04
? PRIOR APPLICATION NUMBER: US 60/207,456
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: US 09/632,366
? PRIOR FILING DATE: 2000-08-03
? PRIOR APPLICATION NUMBER: GB 24263.6
? PRIOR FILING DATE: 2000-10-04
? PRIOR APPLICATION NUMBER: US 60/236,359
? PRIOR FILING DATE: 2000-09-27
? PRIOR APPLICATION NUMBER: PCT/US01/00666
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00664
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00669
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00665
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00668
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00663
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00662
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00661
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: PCT/US01/00670
? PRIOR FILING DATE: 2001-01-30
? PRIOR APPLICATION NUMBER: US 60/234,687
? PRIOR FILING DATE: 2000-09-21
? PRIOR APPLICATION NUMBER: US 09/608,408
? PRIOR FILING DATE: 2000-06-30
? PRIOR APPLICATION NUMBER: US 09/774,203
? PRIOR FILING DATE: 2001-01-29
? NUMBER OF SEQ ID NOS: 49117
? SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
? SEQ ID NO 2772
? LENGTH: 962
? TYPE: DNA
? ORGANISM: Homo sapiens
? FEATURE:
? OTHER INFORMATION: MAP TO AC008125.9
? OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
? OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
? OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
? OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
? OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
? OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
? OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
```





;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 2534  
;; LENGTH: 496  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC011416.1  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6  
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8  
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.5  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
US-09-864-761-2534

Query Match 1.8%; Score 42; DB 10; Length 496;  
Best Local Similarity 52.9%; Pred. No. 0.3;  
Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;  
QY 2038 TGATGATGGGCAATGCCAGTGTGTAGGCCAGAAAGCCACAGTGAAGATCC 2097  
DB 484 TGGTTGGGGGTGGGGAGAGAGCGAGAAAGAGAGAGAGAGAGAGAGAGA 425  
QY 2098 AATAAGAACTTCAAAGCGATGAAGAGCCAGTRAGAGAAACACTGTGAGCCAGAGT 2157  
DB 424 AGAAGAAAGAAAGAAAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAA 365  
QY 2158 CCTCAAGTAGAGCTGAAGAAAGTGAAGCCCAACTCAAGAAAGCAGAG 2207  
DB 364 AGAAGAAAGAAAGAAAGAGAGAGAAAGAAAGAAAGAAAGAAAGAAAG 315

Search completed: November 12, 2002, 15:28:40  
Job time : 101 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 14:47:29 ; Search time 69 Seconds  
(without alignments)  
1537.210 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKNRVS.....KLALLKGNPSSYSKEKIN 796

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_101002:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4163	100.0	796	19 AAM55090	Streptococcus pneu
2	4163	100.0	796	23 ABP54584	S. pneumoniae SP03
3	4163	100.0	819	21 AAB01468	Recombinant varian
4	3218	77.3	821	21 AAB12727	Streptococcus pneu
5	3218	77.3	821	21 AAB12766	Streptococcus pneu
6	3218	77.3	821	23 AAB084026	Truncated variant
7	3218	77.3	840	21 AAB12716	Streptococcus pneu
8	3218	77.3	840	23 AAB075933	Streptococcus pneu
9	2961	71.1	690	21 AAB12745	Streptococcus pneu
10	2961	71.1	690	23 AAB084093	Truncated variant

11	2812	67.5	826	21 AAY91939	S. pneumoniae 92 k
12	2795	67.1	827	21 AAY81662	Streptococcus pneu
13	2788.5	67.0	819	21 AAB01469	Recombinant varian
14	2786	66.9	805	21 AAB12764	Streptococcus pneu
15	2772	66.6	820	21 AAB12755	Streptococcus pneu
16	2772	66.6	838	21 AAB01466	Recombinant varian
17	2771	66.5	807	21 AAB12765	Streptococcus pneu
18	2760	66.3	811	21 AAB12763	Streptococcus pneu
19	2750.5	66.0	819	21 AAB12740	Streptococcus pneu
20	2750.5	66.0	819	21 AAB12754	Streptococcus pneu
21	2750.5	66.0	819	23 AAB04087	Truncated variant
22	2750.5	66.0	838	21 AAB12720	Streptococcus pneu
23	2750.5	66.0	838	23 AAB075934	Streptococcus pneu
24	2733	65.6	834	21 AAB12759	Streptococcus pneu
25	2728	65.5	816	21 AAB12756	Streptococcus pneu
26	2727	65.5	816	21 AAB12757	Streptococcus pneu
27	2720	65.3	811	21 AAB12760	Streptococcus pneu
28	2720	65.3	811	21 AAB12762	Streptococcus pneu
29	2714	65.2	816	21 AAB12758	Streptococcus pneu
30	2713	65.1	811	21 AAB12761	Streptococcus pneu
31	2649.5	63.6	763	19 AAM55095	Streptococcus pneu
32	2649.5	63.6	763	23 ABP54589	S. pneumoniae SP04
33	2431.5	58.4	721	20 AAY05753	Streptococcus pneu
34	2170	52.1	613	21 AAB12730	Streptococcus pneu
35	2170	52.1	613	23 AAB084029	Truncated variant
36	1898	45.6	1126	23 AAB084058	S. pneumoniae deri
37	1891.5	45.4	1365	23 AAB084057	S. pneumoniae deri
38	1888	45.3	555	21 AAB12734	Streptococcus pneu
39	1888	45.3	555	23 AAB084047	Truncated variant
40	1808	43.4	612	23 AAB084031	Truncated variant
41	1795.5	43.1	568	21 AAB12741	Streptococcus pneu
42	1795.5	43.1	568	23 AAB084088	Truncated variant
43	1795.5	43.1	1139	23 AAB084055	S. pneumoniae deri
44	1795.5	43.1	1378	23 AAB084053	S. pneumoniae deri
45	1666	40.0	334	21 AAB12728	Streptococcus pneu

## ALIGNMENTS

RESULT 1			
AAM55090	AAM55090 standard; Protein: 796 AA.		
XX	AAAM55090;		
AC	02-OCT-1998 (first entry)		
DT	Streptococcus pneumoniae SP0036 protein.		
XX			
DE	Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;		
XX	detection; pneumonia; otitis media; meningitis.		
KW	Streptococcus pneumoniae.		
KW	Streptococcus pneumoniae.		
OS	Streptococcus pneumoniae.		
XX			
XX	Key	Location/Qualifiers	
FT	Misc-difference 456	/label= unknown	
FT		/note= "encoded by GNA"	
XX			
PN	W09818930-A2.		
XX			
PD	07-MAY-1998.		
XX			
XX	30-OCT-1997; 97WO-US19422.		
PF			
XX	31-OCT-1996; 96US-0029960.		
PR			
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA			
XX	Choi GH, Hromockyj A, Johnson LS, Kunsch CA;		
FI			
XX	WPI: 1998-272224/24.		
DR			

DR N-PSDB; AANV7351.  
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus  
PT pneumoniae - or their epitope-containing fragments, useful in  
PT protective or therapeutic vaccines, and for diagnosis  
XX  
XX Claim 11; Page 59-60; 118pp; English.  
XX  
CC The present sequence represents a protein from Streptococcus pneumoniae.  
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein  
CC can be useful in vaccines for inducing protective antibodies against  
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.  
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid  
CC are used to detect Streptococcus infection (by usual hybridisation or  
CC amplification methods), also for isolating Streptococcus genes or their  
CC allelic variants. The protein can be used similarly to detect specific  
CC antibodies in standard immunoassays, especially for diagnosing or  
CC monitoring infections. Antibodies which bind the protein are used to  
CC detect corresponding antigens, to purify the protein and for passive  
CC immunisation (optionally coupled to a toxin). Vaccines are administered,  
CC e.g. by injection, orally or through the skin, typically at 0.01-1000  
CC (especially 10-300) mu g/ml per dose.  
XX  
XX

SQ Sequence 796 AA:

Query Match 100.0%; Score 4163; DB 19; Length 796;  
Best Local Similarity 100.0%; Pred. No. 2,7e-299;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGYVT 60  
DB 1 STELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGYVT 60  
OY 61 SHGDHYHYNGKVPYDAIISEELMKDPYKLEDEDIVNEVGKGYIKVYDKGYVYLKDA 120  
DB 61 SHGDHYHYNGKVPYDAIISEELMKDPYKLEDEDIVNEVGKGYIKVYDKGYVYLKDA 120  
OY 121 AHADNVRTKEEINROKQEHSGHREGGTPRNDGAVALARISQGRYTTDDGYIFNADIIEDT 180  
DB 121 AHADNVRTKEEINROKQEHSGHREGGTPRNDGAVALARISQGRYTTDDGYIFNADIIEDT 180  
OY 181 GDAYTYPHGDHYHYIPKNLSASSELAAAFPLSGRNLNSRTYRQNSDNTSRMTWVVS 240  
DB 181 GDAYTYPHGDHYHYIPKNLSASSELAAAFPLSGRNLNSRTYRQNSDNTSRMTWVVS 240  
OY 241 VSNPGTNTNTNSNTNSQASQSDNDISLKLQYKPLPSORHVESDGLVFPDAQTTSRT 300  
DB 241 VSNPGTNTNTNSNTNSQASQSDNDISLKLQYKPLPSORHVESDGLVFPDAQTTSRT 300  
OY 301 ANGVAVPHGDHYHYIPYSOMSELEERIALIPLRYSNHWVPDSRPEQSPQTPPEPSFG 360  
DB 301 ANGVAVPHGDHYHYIPYSOMSELEERIALIPLRYSNHWVPDSRPEQSPQTPPEPSFG 360  
OY 361 POPAPPLKIDSNSLSQSLVRKVGEGYVEEKGISYVFAKOLPSTYVNLSSKLSKQSS 420  
DB 361 POPAPPLKIDSNSLSQSLVRKVGEGYVEEKGISYVFAKOLPSTYVNLSSKLSKQSS 420  
OY 421 VSHHTLAKKENVAPRDOEYDKAYNLLTEAHKALFXNKRNSDFQALDLERLNDSESTN 480  
DB 421 VSHHTLAKKENVAPRDOEYDKAYNLLTEAHKALFXNKRNSDFQALDLERLNDSESTN 480  
OY 481 KEKLVDDLLAFLAPITHTPERLCKPNSQLEYTEDEVRIAO LADKYTTSDDGYIFDEHDIISD 540  
DB 481 KEKLVDDLLAFLAPITHTPERLCKPNSQLEYTEDEVRIAO LADKYTTSDDGYIFDEHDIISD 540  
OY 541 EBDAYVTTPMGSNHWI GKDSLDEKVAQAAYTKKEGILIPSPDADVKANPPGDSAAAY 600  
DB 541 EBDAYVTTPMGSNHWI GKDSLDEKVAQAAYTKKEGILIPSPDADVKANPPGDSAAAY 600  
OY 601 NNKVEKRIPLVRLPYMVEHTVEVNGNLLIPHKDHYHNIRKAWPDDHTYKAPNGYTLED 660  
DB 601 NNKVEKRIPLVRLPYMVEHTVEVNGNLLIPHKDHYHNIRKAWPDDHTYKAPNGYTLED 660

OY 661 LEATTIKYVEHDEPDRPHSDNGNASEHYLGKKHSDPNNKFKADEPVEETPAEPEVP 720  
DB 661 LEATTIKYVEHDEPDRPHSDNGNASEHYLGKKHSDPNNKFKADEPVEETPAEPEVP 720  
OY 721 QVETEKVEAQLKEAVYLLAKYTDSSISKANATETTLAQLRNLTLQIMDNNSIMAEKELLA 780  
DB 721 QVETEKVEAQLKEAVYLLAKYTDSSISKANATETTLAQLRNLTLQIMDNNSIMAEKELLA 780  
OY 781 LKGSNPSVSKEKIN 796  
DB 781 LKGSNPSVSKEKIN 796

RESULT 2

ABP54584 standard; Protein: 796 AA.

ABP54584;

04-SEP-2002 (first entry)

S. pneumoniae SP036 protein sequence SEQ ID NO:56.

Streptococcus pneumoniae; epitope; vaccine; antigenic protein;

antibacterial; Streptococcal infection; detection.

Streptococcus pneumoniae.

US2002061545-A1.

23-MAY-2002.

22-JAN-2001; 2001US-0765272.

30-OCT-1997; 97US-0961083.

(CHOI/) CHOI G. H.

(KUNS/) KUNSCH C. A.

(BARA/) BARASH S. C.

(DILL/) DILLON P. J.

(DOUG/) DOUGHERTY B.

(FANN/) FANNON M. R.

(ROSE/) ROSEN C. A.

Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

Rosen CA;

WPI: 2002-479261/51.

N-PSDB; ABQ84819.

New Streptococcus pneumoniae antigens, useful for detecting

Streptococcus and for preventing or attenuating disease caused by

Streptococcus infection -

Claim 11; Page 27; 70pp; English.

ABQ84792 to ABQ84904 represents nucleic acids which encode the

Streptococcus pneumoniae antigens given in ABP54557 to ABP54659.

The S. pneumoniae antigens have antibacterial activity and can be

used in vaccines. The S. pneumoniae antigens can also be used to

prevent or attenuate a Streptococcal infection in an animal. The

CC polynucleotides encoding the S. pneumoniae antigens can be used to

CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent

CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)

CC which are used in an example from the present invention.

Sequence 796 AA;

Query Match 100.0%; Score 4163; DB 23; Length 796;

Best Local Similarity 100.0%; Pred. No. 2,7e-299;

Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYELGLYQARTVKENNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGYVT 60



|||||  
Db 1 SYELGLYQARYKENNRYSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYT 60  
OY 61 SHGDHYHYNGKVPYDAIISELLMKDPYKIKDEDIYVEVGGYVIKDGKYYVYLKDA 120  
Db 61 SHGDHYHYNGKVPYDAIISELLMKDPYKIKDEDIYVEVGGYVIKDGKYYVYLKDA 120  
OY 121 AHADVNRTEEINROKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180  
Db 121 AHADVNRTEEINROKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180  
OY 181 GDAYIVPHGDHYHYIPKNELASASELAAPAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPS 240  
Db 181 GDAYIVPHGDHYHYIPKNELASASELAAPAFLSGRGNLSNSRTYRRQNSDNTSRTNWVPS 240  
OY 241 VSNPGTNTNTNSNTNSQASQSDNDISLLKQYKPLSQGHVESDGLVFPDPAQITST 300  
Db 241 VSNPGTNTNTNSNTNSQASQSDNDISLLKQYKPLSQGHVESDGLVFPDPAQITST 300  
OY 301 ARGVAVPHGDHYHYIPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQTPPEPSFG 360  
Db 301 ARGVAVPHGDHYHYIPYQMSSELEERARIIPLRYSNHWVDSRPEQSPQTPPEPSFG 360  
OY 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVFAKDLPSETYKNLESKLSKQES 420  
Db 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVFAKDLPSETYKNLESKLSKQES 420  
OY 421 VSHHTLAKKENVAPRDOEFYDKAYNLTPFAHKALFXKNGRNSDFQALDKLERLNDSTN 480  
Db 421 VSHHTLAKKENVAPRDOEFYDKAYNLTPFAHKALFXKNGRNSDFQALDKLERLNDSTN 480  
OY 481 KEKLVDDLAFAPITTHPERLGKPNQSOLEYTEDEYVRIQALADKYTSDGYIFDEHDIISD 540  
Db 481 KEKLVDDLAFAPITTHPERLGKPNQSOLEYTEDEYVRIQALADKYTSDGYIFDEHDIISD 540  
OY 541 EEDDAYVTTPMGSHWIGKDSLSDKERVAQAQYTKKGIIPSPDADVKANPTGDSAAIY 600  
Db 541 EEDDAYVTTPMGSHWIGKDSLSDKERVAQAQYTKKGIIPSPDADVKANPTGDSAAIY 600  
OY 601 NVKGEKRIPLVRLPYMETHVEVNGNLIIPKDHYNHIFKAWDDHYYKAPNGYTTLED 660  
Db 601 NVKGEKRIPLVRLPYMETHVEVNGNLIIPKDHYNHIFKAWDDHYYKAPNGYTTLED 660  
OY 661 LEFATIKYVEHPDERPHSNDGWSNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 720  
Db 661 LEFATIKYVEHPDERPHSNDGWSNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 720  
OY 721 QVETEKVEAQLEAEVLLAKYTDSSILKANATELAGLRNNLTQIMDNNSINAAEKLLA 780  
Db 721 QVETEKVEAQLEAEVLLAKYTDSSILKANATELAGLRNNLTQIMDNNSINAAEKLLA 780  
OY 781 LKGSNPSSVSKEKIN 796  
Db 781 LKGSNPSSVSKEKIN 796

RESULT 3

ID AAB01468 standard: Protein: 819 AA.

XX AAB01468:

XX 20-OCT-2000 (first entry)

XX Recombinant variant of Sp36 (Sp36a) of S. pneumoniae.

XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;

XX histidine triad residue; Sp36; antibody; otitis media;

XX nasopharyngeal infection; bronchial infection; bronchitis; sepsis;

XX meningitis; lobar pneumonia.

XX Streptococcus pneumoniae.

OS

FH Key Location/Qualifiers  
FT Region 63..68  
FT /label= Histidine triad residue  
FT Region 118..145  
FT /label= Coiled coil region  
FT Region 189..194  
FT /label= Histidine triad residue  
FT Region 309..314  
FT /label= Histidine triad residue  
FT Region 406..434  
FT /label= Coiled coil region  
FT Region 462..493  
FT /label= Coiled coil region  
FT Region 550..555  
FT /label= Histidine triad residue  
FT Region 634..639  
FT /label= Histidine triad residue  
FT Region 724..751  
FT /label= Coiled coil region

W0200037105-A2.

29-JUN-2000.

21-DEC-1999; 99WO-US30390.

21-DEC-1998; 98US-0113048.

(MEDI-) MEDIMUNE INC.

Johnson LS, Koenig S, Adamou JE;

WPI: 2000-452129/39.

N-PSDB: AAA47604.

Vaccine useful for prophylaxis and treatment of pneumococcal infections

such as otitis media, nasopharyngeal and bronchial infections,

comprises Streptococcus pneumoniae proteins

Claim 1: Page 61-64; 70pp; English.

Although a number of proteins have been suggested as being involved in the pathogenicity of Streptococcus pneumoniae, there still remains a need to identify polypeptides having epitopes in common from various strains of S. pneumoniae in order to utilise such polypeptides in vaccines to protect against a wide variety of S. pneumoniae. New vaccine compositions are described which comprise a Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino acids in length that comprise at least one histidine triad residue (HxxHxx) or a coiled-coil region, or an antibody directed against these features. The vaccine is useful in protecting against infection by Streptococcus pneumoniae. The vaccine composition comprising antibodies to is useful for passive immunization for treating pneumococcal infections which includes otitis media, nasopharyngeal and bronchial infections.

Sequence 819 AA:

Query Match 100.0%; Score 4163; DB 21; Length 819;  
Best Local Similarity 99.9%; Pred. No. 2.9e-299;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYELGLYQARYKENNRYSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYT 60  
Db 21 SYELGLYQARYKENNRYSYIDGKATQKTENLTPEVSKREGINAEQIVIKITDQGYT 80  
OY 61 SHGDHYHYNGKVPYDAIISELLMKDPYKIKDEDIYVEVGGYVIKDGKYYVYLKDA 120  
Db 81 SHGDHYHYNGKVPYDAIISELLMKDPYKIKDEDIYVEVGGYVIKDGKYYVYLKDA 140  
OY 121 AHADVNRTEEINROKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 180  
Db 141 AHADVNRTEEINROKQEHSGHREGGTPRNDGAVALARSGQRYTTDDGYIFNASDIIEDT 200

```
QY 181 GDAYIYPHGDHYHYIPKNELSASELAEEAFLSGRGNLSNRTYRONSNDNSTRNWPVS 240
    |||||||
Db 201 GDAYIYPHGDHYHYIRKNELSASELAEEAFLSGRGNLSNRTYRONSNDNSTRNWPVS 260
QY 241 VSNPGTTNTNTSNNSNTNSQASQSNDSIDSLKQLYKPLSQRHVESDGLVEFDAQITSRT 300
    |||||||
Db 261 VSNPGTTNTNTSNNSNTNSQASQSNDSIDSLKQLYKPLSQRHVESDGLVEFDAQITSRT 320
QY 301 ARGVAAPPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQPTPEPSPG 360
    |||||||
Db 321 ARGVAAPPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQPTPEPSPG 380
QY 361 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVEAKDLPSETVKNLESKLSKQES 420
    |||||||
Db 381 POPAPNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVEAKDLPSETVKNLESKLSKQES 440
QY 421 VSHITLAKKENVAPRDOEYDKAYNLLTEAHKALFYKNGNSDFOALDKLERLNDESTN 480
    |||||||
Db 441 VSHITLAKKENVAPRDOEYDKAYNLLTEAHKALFYKNGNSDFOALDKLERLNDESTN 500
QY 481 KEPLVDDLAFIAPLITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIEPDHIIISD 540
    |||||||
Db 501 KEPLVDDLAFIAPLITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIEPDHIIISD 560
QY 541 EGDAYYTPHMGSHMWIGKDSLSDKEKYAAQAYTKEKGILPPSPDADYKANPTGDSAAATY 600
    |||||||
Db 561 EGDAYYTPHMGSHMWIGKDSLSDKEKYAAQAYTKEKGILPPSPDADYKANPTGDSAAATY 620
QY 601 NRVGKGRKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGYTLDE 660
    |||||||
Db 621 NRVGKGRKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGYTLDE 680
QY 661 LEFTIKYYVEHPDERHSNDGMGNASEHVLGKDHSEDPKKNKADBEVEETPAPEVP 720
    |||||||
Db 681 LEFTIKYYVEHPDERHSNDGMGNASEHVLGKDHSEDPKKNKADBEVEETPAPEVP 740
QY 721 OVETEKEVQAOLKEAEVLLAKVTDSSILKANATEYIAGLRNNLTQIMDNNSIMAEAKELLA 780
    |||||||
Db 741 OVETEKEVQAOLKEAEVLLAKVTDSSILKANATEYIAGLRNNLTQIMDNNSIMAEAKELLA 800
QY 781 LKGSNPPSSYSKEKIN 796
    |||||||
Db 801 LKGSNPPSSYSKEKIN 816

RESULT 4
AAB12727
ID AAB12727 standard; Protein; 821 AA.
XX
AC AAB12727;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
```

```
XX
DR MPI: 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 25; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-11M protein antigen.
XX
SQ Sequence 821 AA:
XX
Query Match 77.3%; Score 3218; DB 21; Length 821;
Best Local Similarity 75.2%; Pred. No. 2,8e-229;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
QY 1 SYELGLYQARTYKENNRVSYIDGKQATOKTENLTPEVSKRGINAEOIVIKITPOGYVT 60
    :|||||:|||||
Db 2 AYELGLHQQYTKENNRVSYIDGKQATOKTENLTPEVSKRGINAEOIVITPDGYVT 61
QY 61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLDIEDIVNEVGKGYIVKGYKYYYLDA 120
    |||||||
Db 62 SHGDHYHYNGKVPYDAIISEELMKDPNYQQLKDSIDYNEIKGYIVKGYKYYYLDA 121
QY 121 AHADNVRKEELINRQKQESHQREGTFRNDGAVALARSQGYTTDDGYIFNADSIIDET 180
    |||||||
Db 122 AHADNVRKEELINRQKQESHQREGTSANDGAVAFARQGYTTDDGYIFNADSIIDET 181
QY 181 GDAYIYPHGDHYHYIPKNELSASELAEEAFLSGRGNLSNRTYRONSNDNSTRNWPVS 240
    |||||||
Db 182 GDAYIYPHGDHYHYIPKNELSASELAEEAFLSGRGNLSNRTYRONSNDNSTRNWPVS 241.
QY 241 VSNPGTTNTNTSNNSNTNSQASQSNDSIDSLKQLYKPLSQRHVESDGLVEFDAQITSRT 300
    |||||||
Db 242 VSNPGTTNTNTSNNSNTNSQASQSNDSIDSLKQLYKPLSQRHVESDGLVEFDAQITSRT 301
QY 301 ARGVAAPPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQPTPEPSPG 360
    |||||||
Db 302 ARGVAAPPHGHHYHFIPYSQMSLEERARIIPLRYSNMHWPPSRPQSPQPTPEPSPG 361
QY 361 POPAPNLK-IDSN--SSLVSQLVRKVGEGYVEEKGISRYVEAKDLPSETVKNLESKLS 416
    |||||||
Db 362 POPAPNLK-IDSN--SSLVSQLVRKVGEGYVEEKGISRYVEAKDLPSETVKNLESKLS 421
QY 417 KOESVSHITLAKKENVAPRDOEYDKAYNLLTEAHKALFYKNGNSDFOALDKLERLN 476
    |||||||
Db 422 KOESVSHITLAKKENVAPRDOEYDKAYNLLTEAHKALFYKNGNSDFOALDKLERLN 481
QY 477 ESTNNKELVDDLAFIAPLITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIEPDH 536
    |||||||
Db 482 VSDSKVCLVDDLAFIAPLITHPERLGKPNQOIEYTEDEVRILAQADKTTSDGIIEPDH 541
QY 537 IISDGDAYVTPHMGSHMWIGKDSLSDKEKYAAQAYTKEKGILPPSPDADYKANPTGDS 596
    |||||||
Db 542 IISDGDAYVTPHMGSHMWIGKDSLSDKEKYAAQAYTKEKGILPPSPDADYKANPTGDS 601
QY 597 AAIYNRKVGKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGY 656
    |||||||
Db 602 AAIYNRKVGKRIPLVRLPYVVEHTVEVKNGNLIIPKRDYHNIRKFAFPDHTYKAPNGY 661
QY 657 TLEDLEFATIKYYVEHPDERHSNDGMGNASEHVLGKDHSEDPKKNKADBEVEETPA 706
    |||||||
Db 662 TLEDLEFATIKYYVEHPDERHSNDGMGNASEHVLGKDHSEDPKKNKADBEVEETPA 721
QY 707 -----EEPEETPAPEVPQVETEKEVAOLKEAEVLLAKVTDSSL 746
```



PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and  
PT epitope-bearing polypeptides, useful as vaccine components for treating  
PT or preventing streptococcal infections such as otitis media,  
PT meningitis, and bacteraemia

XX Example 1; Page -; 113pp; English.

CC The invention describes an isolated polypeptide (I) with 70-90%  
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of  
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)  
CC comprising (I) is useful for therapeutic or prophylactic treatment of  
CC meningitis, otitis media, bacteraemia or pneumonia infection in an  
CC individual susceptible to these disorders. (II) is also useful for  
CC therapeutic or prophylactic treatment of any streptococcal bacterial  
CC infection (e.g., caused by Streptococcus pneumoniae, group A  
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such  
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novae or  
CC Staphylococcus aureus) in an individual susceptible to the infection.  
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation  
CC techniques. The Streptococcus polypeptides are useful in a diagnostic  
CC test for S. pneumoniae infection. (III) is useful for designing DNA  
CC probes for use in detecting the presence of Streptococcus in a biological  
CC sample suspected of containing the bacteria. The DNA probes may also be  
CC used for detecting circulating S. pneumonia nucleic acid in a sample for  
CC diagnosing streptococcal infections. This sequence represents a truncate  
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,  
CC Note: This sequence does not appear in the specification but has  
CC been created according to information given in the invention.

XX Sequence 821 AA:

Query Match 77.3%; Score 3218; DB 23; Length 821;

Best Local Similarity 75.2%; Pred. No. 2.8e-229;  
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKNRRVSYIDGKATQKTENTLTPDEVSKREGINAEQIVIKITDGGYVT 60  
DB 2 AVELGLHQATVKNRRVSYIDGKATQKTENTLTPDEVSKREGINAEQIVIKITDGGYVT 61  
QY 61 SHGDHHYNGKPYDAIISEELMKDPNYKLDKEDIVNEKGGYIKYDKGKYYVLKDA 120  
DB 62 SHGDHHYNGKPYDAIISEELMKDPNYKLDKEDIVNEKGGYIKYDKGKYYVLKDA 121  
QY 121 AHADNVRTEKEIRKQESQHQHREGTTPRNDGVALARSGRTTDDGYIFNANSDIIEPT 180  
DB 122 AHADNVRTEKEIRKQESQHQHREGTTPRNDGVALARSGRTTDDGYIFNANSDIIEPT 181  
QY 181 GDAYIVPHGDHYHYIPKNELSASELAEEFLSGRGNLSNRTYRONSNDNTSRFMWVPS 240  
DB 182 GDAYIVPHGDHYHYIPKNELSASELAEEFLSGRGNLSNRTYRONSNDNTSRFMWVPS 241  
QY 241 VSPGTTNTNTSNSTNSQASQNDISLKLQYKLPISQRYVESDGLVFPDPAQITTSRT 300  
DB 242 VSPGTTNTNTSNSTNSQASQNDISLKLQYKLPISQRYVESDGLVFPDPAQITTSRT 301  
QY 301 AGGVAAPHGDHYHYIPYQSMSELEERIALIPLRYRSNMWVPSPREQSPOTPEPSPG 360  
DB 302 AGGVAAPHGDHYHYIPYQSMSELEERIALIPLRYRSNMWVPSPREQSPOTPEPSPG 361  
QY 361 POPAPMLK-IDSN--SLVSQLVKRVGEGYVEEKGISRYEPAKDLPELVKNELSKLS 416  
DB 362 POPAPMLK-IDSN--SLVSQLVKRVGEGYVEEKGISRYEPAKDLPELVKNELSKLS 421  
QY 417 KQESVSHITLAKKENVAPPDQEFYDKAVNLTFAHKAFLXNKGKNSDFQALDKLERLND 476  
DB 422 KQESVSHITLAKKENVAPPDQEFYDKAVNLTFAHKAFLXNKGKNSDFQALDKLERLND 481  
QY 477 ESNKKEKLVDDLALAPLTHPERLCKPNSQTEYTEDVRIQLADKYTTSQYIDEHD 536  
DB 482 VSSDKKLVDDLALAPLTHPERLCKPNSQTEYTEDVRIQLADKYTTSQYIDEHD 541  
QY 537 IISDEGDVAVTTHMGSHWIGKDSLSDEKVAQAQATKKGILPSPDADVANKPITGSDA 596

DB 542 IISDEGDVAVTTHMGSHWIGKDSLSDEKVAQAQATKKGILPSPDADVANKPITGSDA 601  
QY 597 AATYRVKGEKRIPLRYLPMYEHVYKNGNLIIPKHCHYINIKFAMDDHTTAPKPY 656  
DB 602 EATYRVKGEKRIPLRYLPMYEHVYKNGNLIIPKHCHYINIKFAMDDHTTAPKPY 661  
QY 657 TIEDLATIKYVEHNDENPHSDNGFGNSDHYQRKKNQADPTNTEKSEKPOTEKE 706  
DB 662 TIEDLATIKYVEHNDENPHSDNGFGNSDHYQRKKNQADPTNTEKSEKPOTEKE 721  
QY 707 -----EEPEETPAPEVPOVEYERKVAQKAEVLLAKYVDSGL 746  
DB 722 EETPREKPOSEKPEKPTPEEPEESPESEPEQYETEKEVKELREAPDLGKIQDPII 781  
QY 747 KANATETLAGLRNNLTQLQIMDNNSIMAEKLLALLKGS 785  
DB 782 KSNAKETTLGLKNNLTLFGQDNNTIWAEEKLLALLKES 820

RESULT 7

ID AAB12716 standard; Protein: 840 AA.

AC AAB12716;

XX 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-11 protein antigen SEQ ID NO:4.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KW otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

XX WO200039299-A2.

XX PD 06-JUL-2000.

XX PF 20-DEC-1999; 99WO-CA01218.

XX PR 23-DEC-1998; 98US-0113800.

XX PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

DR N-PSDB; AAA65731.

XX Claim 18; Fig 4; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens  
CC have bactericidal activity. The nucleic acids, encoding the protein  
CC antigens, may be used for the recombinant production of the proteins  
CC they encode. The protein antigens may then be used as vaccines for the  
CC prevention and treatment of Streptococcal infections in mammals  
CC (especially humans) which result in, e.g. meningitis, otitis media,  
CC bacteraemia and/or pneumonia. The present sequence represents the  
CC S. pneumoniae BVH-11 protein antigen.

XX Sequence 840 AA:

Query Match 77.3%; Score 3218; DB 21; Length 840;

Best Local Similarity 75.2%; Pred. No. 2.9e-229;  
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;

QY 1 SYELGLYQARTVKNRRVSYIDGKATQKTENTLTPDEVSKREGINAEQIVIKITDGGYVT 60

```

Db      21 AVELGLHOAQTVKENNRVSYIDGKQATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT
OY      61 SHGDHYHYNGKVPYDAIISELLMKDPYKDKDDIYNEVGGVIVYKDGKYYVYLKDA 120
Db      81 SHGDHYHYNGKVPYDAIISELLMKDPYKDKDDIYNEVGGVIVYKDGKYYVYLKDA 140
OY      121 AHADNVRTKEEINROKOEHSQHREGTTPNDGAVALARSOGRYTTDDGYIFNADSIIEDT 180
Db      141 AHADNVRTKEEINROKOEHSQHREGTSANDGAVAFARSOGRYTTDDGYIFNADSIIEDT 200
OY      181 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRONSNDNTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRONSNDNTSRTNWVPS 260
OY      241 VSNPGTTNTNTNSNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAITST 300
Db      261 VSNPGTTNTNTNSNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAITST 320
OY      301 ARGVAVPHGDHYHYIPYSQMSLEERIRIIPLRYSNMHWVDSRPEQSPQPTPEPSPG 360
Db      321 ARGVAVPHGNHYHFIPYEGMSLEKRIARIIPLRYSNMHWVDSRPEQSPQPTPEPSPG 380
OY      361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYVEEKGISRYVFAFDLDBSETVKNLESKUS 416
Db      381 POPAPNPQAPSNPIDEKLVKCAVVRKVGCGYVEENGVSRYIPAKNLSAETAGIDSKLA 440
OY      417 KOESVSHLTAKKENVAPRDOEFYKAVNLLTEAKKALFXNKGSRSDQALDKLEIRLND 476
Db      441 KOESLSHKLGAACKTDLPSDRREFYKAVDYLARIHODLDLNDNGROVDEALDNLIERLND 500
OY      477 ESTNKEKLVDDLAFAPITPHERLGKPNQSQIETEYDEVERIAQLADKYTSDGYTFDEHD 536
Db      501 VSSDKVKLVDDLAFAPITPHERLGKPNQAQITTYTDDDEIQVAKLAGKYTEDEGYTFDPRD 560
OY      537 IISDGDAYIVPHMGSHWIGKDSLSDEKVAQAQAYTEKGLPPSPDADVANKPTGDSA 596
Db      561 IISDGDADAYIVPHMTHSHIKKDSLSAEARAQAQAYAKKEGILPSTJHODSGNTEAKGA 620
OY      597 AAIYVRKVEKRIPLVRLPYVWEHVEYVNGNLIIIPKHQYINIFPAMDHDHYTAPANGY 656
Db      621 EAIYVRKAKKAVPLDRMPYNLQITVEYKNGSLIIPHDHYHNIFEMWDEGLTEAPKGY 680
OY      657 TLEDLFATIKYVVEPDERPHSNDGMSNASEHVLGKDKHSEDPNKNFKAD----- 706
Db      681 TLEDLATYKYVVEHPNERPHSDNGFGNASHYQNRKNQADPTNQTEKPESEKPOTEKPE 740
OY      707 -----EPPVEETPAPEVPOVETEKVDAQLKEAVYLLAKYTDSSL 746
Db      741 EETPREEKPOSEKPEPKPTTEPEEESPESEBPOVETEKVEKILREADLLGKIODPII 800
OY      747 KANATELLAGLNNNTLTQIMONNSIMAEKLLALLKGS 785
Db      801 KSNAKETTLGLKNNLLFTQDNNNTIMAEKLLALLKES 839

RESULT 8
AAU75933
ID      AAU75933 standard; Protein: 840 AA.
XX      AAU75933;
AC      08-MAY-2002 (first entry)
DE      Streptococcus pneumoniae BVH-11 protein.
XX      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW      pneumonia; streptococcal bacterial infection.
XX      Streptococcus pneumoniae.
OS      WO200198334-A2.
PN      XX

```

```

PD      27-DEC-2001.
XX      19-JUN-2001; 2001WO-CA00908.
PF      20-JUN-2000; 2000US-212683P.
PR      (SHIR-) SHIRE BIOCHEM INC.
XX      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      N-PSDB; ABK15103.
DR      WPI: 2002-122272/16.
DR      N-PSDB; ABK15103.
XX      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT      epitope-bearing polypeptides, useful as vaccine components for treating
PT      or preventing streptococcal infections such as otitis media,
XX      meningitis, and bacteraemia
PS      Example 1; Fig 7; 113pp; English.
XX      The invention describes an isolated polypeptide (I) with 70-90%
CC      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC      comprising (I) is useful for therapeutic or prophylactic treatment of
CC      meningitis, otitis media, bacteraemia or pneumonia infection in an
CC      individual susceptible to these disorders. (II) is also useful for
CC      therapeutic or prophylactic treatment of any streptococcal bacterial
CC      infection (e.g., caused by Streptococcus pneumoniae, group A
CC      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
CC      Streptococcus aureus) in an individual susceptible to the infection.
CC      A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC      techniques. The Streptococcus polypeptides are useful in a diagnostic
CC      test for S. pneumoniae infection. (III) is useful for designing DNA
CC      probes for use in detecting the presence of Streptococcus in a biological
CC      sample suspected of containing the bacteria. The DNA probes may also be
CC      used for detecting circulating S. pneumonia nucleic acid in a sample for
CC      diagnosing streptococcal infections. This is the amino acid sequence of
CC      Streptococcus pneumoniae protein BVH-11, used to create the antigenic
CC      peptides described in the method of the invention.
SQ      Sequence 840 AA:
      77.3%; Score 3218; DB 23; Length 840;
Query Match 75.2%; Pred. No. 2,9e-229;
Best Local Similarity 65; Mismatches 104; Indels 34; Gaps 3;
Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
OY      1 SYELGLYQARTVKENNRVSYIDGKQATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT 60
Db      21 AVELGLHOAQTVKENNRVSYIDGKQATOKTENLTPDEVSKREGINAEOIVIKITDOGYVT 80
OY      61 SHGDHYHYNGKVPYDAIISELLMKDPYKDKDDIYNEVGGVIVYKDGKYYVYLKDA 120
Db      81 SHGDHYHYNGKVPYDAIISELLMKDPYKDKDDIYNEVGGVIVYKDGKYYVYLKDA 140
OY      121 AHADNVRTKEEINROKOEHSQHREGTTPNDGAVALARSOGRYTTDDGYIFNADSIIEDT 180
Db      141 AHADNVRTKEEINROKOEHSQHREGTSANDGAVAFARSOGRYTTDDGYIFNADSIIEDT 200
OY      181 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRONSNDNTSRTNWVPS 240
Db      201 GDAYIVPHGDHYHYIPKNELSASELAAEAFISGRGNLSNSRTYRONSNDNTSRTNWVPS 260
OY      241 VSNPGTTNTNTNSNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAITST 300
Db      261 VSNPGTTNTNTNSNSNTNSQASQSDNDISLKLQYKPLPSQHHVSDGLIFPDAITST 320
OY      301 ARGVAVPHGDHYHYIPYSQMSLEERIRIIPLRYSNMHWVDSRPEQSPQPTPEPSPG 360
Db      321 ARGVAVPHGNHYHFIPYEGMSLEKRIARIIPLRYSNMHWVDSRPEQSPQPTPEPSPG 380
OY      361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYVEEKGISRYVFAFDLDBSETVKNLESKUS 416

```

Db	361	POPAPNPQPAASNPIDKELVKEARVKGDSYVEEENGVSRYIPAKNLSAETAGIDSKLA	440
Oy	417	KOESVSHTLTAKKENVAPRDQEFYDKAYNLTEAHKALFYNKGRNSDFQALDKLIERLND	476
Db	441	KOESLSHKLGAKKTDLPSSDREFYNKAYDLLARIHODLLDNKGRQVDFEALDNLLERLKD	500
Oy	477	ESTNKERLYVDLLAFAPITHPERBLCKPNSQIETDEVERIAQLADKYTSDGYTFDEHD	536
Db	501	VSSDVKYLVDDILAFAPIRHPERLCKPNAQITYTDEDEIQVAKLACKYTTEDGYTFDPDD	560
Oy	537	IISDEGDAYVYPHAGSHHWIGKSDLSDBKEVAAQAYATKEKGLPSPDPADVANKPTGDSA	596
Db	561	ITSDEGDAYVYPHFTSHHWIKKSDLSAEBAQAQAAKAKGKGLPSTDHQDSGNTEAAGA	620
Oy	597	AAIYNRVKGERKIDLVRLPYVVEHTEVVKNGNLIIPKHQYHNIRKFAWFDDHTFYKAPGCV	656
Db	621	EAIYNRVKAAKVPDLRMPNQLQATVEVKNKGSLLIPHYDHNIRKFEWFDDECLYEAAPGY	680
Oy	657	TLEDLFAIRIKYVYHPDEPRHSNDGNGNASEHVLGKKHSDSEPNKRFKD-----	706
Db	681	TLEDLFAIVKYYVEHPNERPDSNGFENASDHQARKNKQAQADTNQTEKPESEKPOTERKE	740
Oy	707	-----EELPEETPAPEVPOVETEKEVSEAOLKEAFVLLAKYTDSSL	746
Db	741	EETPREEKPOSEKPESEKPRPEEPPEESPESEPOVETEKVEBKLRADDLGKIQDPII	800
Oy	747	KANATETLAGLRNNLTLOINDNNSINAAEKKLLALLKGS	785
Db	801	KSNAKETLTGLKNNLTFGQDNNNTIMAEKLLALLKES	839
RESULT 9			
AAB12745	AAB12745 standard; Protein; 690 AA.		
XX	AAB12745:		
DT	23-NOV-2000 (first entry)		
XX	Streptococcus pneumoniae NEW16 protein antigen SEQ ID NO:79.		
DE	.		
KW	Streptococcus pneumoniae: BWH-3; BWH-11; BWH-28; antigen; vaccine; prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia; otitis media; pneumonia; immunisation; bactericidal.		
OS	Streptococcus pneumoniae.		
PN	WO200039299-A2.		
PD	06-JUL-2000.		
XX	20-DEC-1999; 99WO-CA01218.		
PF	23-DEC-1998; 98US-0113800.		
PR	(BIOC-) BIOCHEM PHARMA INC.		
PA	Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;		
PI	WPI: 2000-452397/39.		
DR	Claim 18; Fig 44; 106pp; English.		
XX	The present invention describes nucleic acids (I) encoding protein antigens (II) from Streptococcus pneumoniae. The protein antigens have bactericidal activity. The nucleic acids, encoding the protein antigens, may be used for the recombinant production of the proteins they encode. The protein antigens may then be used as vaccines for the prevention and treatment of Streptococcal infections in mammals (especially humans) which result in, e.g. meningitis, otitis media,		

[illegible]









```

Db 307 TARGVAVPHGHNHFIPIYQOMSELEKRIARIILPLRYRSNHWVDSRPQSPQSTPEPSP 366
QY 360 GPQAPNLK-IDSN---SLVSQLRVKGEVYEEKEKISRVRFAKDLPSERVKNLESL 415
Db 367 SQQPPNPQAPASNPIDIEKLYEAVKVGDDGVFEENGVSRIIPAKDLISAETPAAGIDSL 426
QY 416 SKQESVSHLPFAKKENVAVRDQEFYDAKAVNLLTEAHKALFXNKGNSDFQALDLERLN 475
Db 427 AKQESLSHKLGAKKTDLPSDSREFYNKAYDLARIHODLLDKKGQVDFEALDNLLERLK 486
QY 476 DESTKEKLVLDLAFAPITPERLGRPNQSIETEDVARIQADLYTSDGIIFDEH 535
Db 487 DVPSSKAVLYDDILAFAPIRHPERLGRNAQITTDDEIOYAKLAGYTTEDGYIFDPR 546
QY 536 DIISDEGAYVTPPHGSHWIGKDSLSDKEKVAQAQYTRKEGILPSPDADYKANPTGDS 595
Db 547 DITSDEGAYVTPPHMTHSHWIKKDSLSEKRAAQAQYAKKGLTPSPDHDQDSCTEAKG 606
QY 596 AAATYRVKGEKRIPLVRLPYVEHTVEVKNGLIIPKDHYNIKFAMFDDHTYKAPNG 655
Db 607 ABAIYNRYAKKAVPLDRMPYNLQYTVENKNGSLIIPHYDHNLIKFEWDEGLYEAAPKG 666
QY 656 YLLEDELATIKYVYEHPRPHSNDGKNASHEVYLGKDHSDPKNFKADEE-----P 709
Db 667 YLLEDELATIKYVYEHPRPHSDNGFGNASDHYORNNKGADTNQTEKPESEKPOTEKP 726
QY 710 VEET-----PAPEPVQVETEKVPAOLKEAEVLLAKYTDSL 746
Db 727 EETPREKPESEKPESEKPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 786
QY 747 KANATETLAGLRNNLTLOIMDNNSIMAEKLLALLKGS 785
Db 787 KSNAKETLGLKNNLFLFGQDNNTIMAEKLLALLKGS 825

RESULT 13
AAB01469
ID AAB01469 standard; Protein: 819 AA.
XX
AC AAB01469;
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 (Sp36B) of S. pneumoniae.
XX
KW Streptococcus pneumoniae; infection; vaccine; colled coll region;
KW histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia.
XX
OS Streptococcus pneumoniae.
XX
RN WO200037105-A2.
XX
PD 29-JUN-2000.
XX
PF 21-DEC-1999; 99WO-US30390.
XX
PR 21-DEC-1998; 98US-0113048.
PA (MEDI-) MEDIMUNE INC.
XX
PI Johnson LS, Koenig S, Adamou JE;
XX
DR WPI: 2000-452129/39.
DR N-PSDB: AAA47605.
XX
PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections,
PT comprises Streptococcus pneumoniae proteins
XX
PS Claim 1; Page 65-69; 70pp; English.
XX

```

```

CC Although a number of proteins have been suggested as being involved
CC in the pathogenicity of Streptococcus pneumoniae, there still remains
CC a need to identify polypeptides having epitopes in common from
CC various strains of S. pneumoniae in order to utilize such
CC polypeptides in vaccines to protect against a wide variety of
CC S. pneumoniae. New vaccine compositions are described which comprise a
CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
CC acids in length that comprise at least one histidine triad residue
CC (HxxHx) or a colled-coll region, or an antibody directed against
CC these features. The vaccine is useful in protecting against infection
CC by Streptococcus pneumoniae. The vaccine composition comprising
CC antibodies to is useful for passive immunization for treating
CC pneumococcal infections which includes otitis media, nasopharyngeal
CC and bronchial infections.
XX
S0 Sequence 819 AA:
Query Match 67.0%; Score 2788.5; DB 21; Length 819;
Best Local Similarity 66.7%; Pred. No. 1.8e-197;
Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;
QY 1 SYELGLYQA-RIVKNNRVSITDQKATQKTENLTPDEYSKREGINAQIVYKITDQGV 59
Db 21 SYELGRYQAGQDKKSNRYAYIDGQAGAKENLTPDEYSKREGINABQIVYKITDQGV 80
QY 60 TSHGDHYHYNKGKVPYDAIISELLMKDPNFKLDEIVNEKGGYVIRKDGKYYVYLKD 119
Db 81 TSHGDHYHYNKGKVPYDAIISELLMKDPNFKLDEIVNEKGGYVIRKDGKYYVYLKD 140
QY 120 AAHADNVTKKEIRINQKQESHOHREGGTPRNDGAVALARSQGRYTTDGYIFNASDIED 179.
Db 141 AAHADNITKKEIRKQKQKERSHNHS---RADNVAARAAQGRYTTDGYIFNASDIED 197
QY 180 TGDATVPHGCHYHYIPKNEISASLAAAEAFLSGRGULNSRTRYRONSNTSTKNVP 239
Db 198 TGDATVPHGCHYHYIPKNEISASLAAAEAFYWG-----KQGRSSSSSYNA 246
QY 240 SVSNNGTNTNTNNSNSQSNDSIDSLKQYKPLPSORHESGLVDPQOTSR 299
Db 247 NPAQRLSENHLTYTPHYQ--NOGENISSLRELYAKPLSRHYESGLIFDPQOTSR 305
QY 300 TARGVAVPHGCHYHYIPYQOMSELEKRIARIILPLRYRSNHWVDSRPQSPQSTPEPSP 359
Db 306 TARGVAVPHGHNHFIPIYQOMSELEKRIARIILPLRYRSNHWVDSRPQSPQSTPEPSP 365
QY 360 GPQAPNLKIDSNSSLVQVAKVGEVYEEKEKISRVRFAKDLPSERVKNLESLKQ 419
Db 366 SQQAPASNPID--GRLVEAVKVGDDGVFEENGVSRIIPAKDLISAETPAAGIDSLKQ 423
QY 420 SVSHTLTFAKKENVAVRDQEFYDAKAVNLLTEAHKALFXNKGNSDFQALDLERLNDEST 479
Db 424 SLSHKLGTKTKTDLPSDSREFYNKAYDLARIHODLLDKKGQVDFEALDNLLERLKDVSS 483
QY 480 NKEKLVLDLAFAPITPERLGRPNQSIETEDVARIQADLYTSDGIIFDEHDIIS 539
Db 484 DKVRLVEDILAFAPITPERLGRPNQSIETEDVARIQADLYTSDGIIFDEHDIIS 543
QY 540 DEGDAYVTPPHGSHWIGKDSLSDKEKVAQAQYTRKEGILPSPDADYKANPTGSAAT 599
Db 544 DEGDAYVTPPHMTHSHWIKKDSLSEKRAAQAQYAEKGLTPSPDHDQDSCTEAKGAAEI 603
QY 600 YNRVYGEKRIPLVRLPYVEHTVEVKNGLIIPKDHYNIKFAMFDDHTYKAPGYTLE 659
Db 604 YNRVYAAKAVPLDRMPYNLQYTVENKNGSLIIPHYDHNLIKFEWDEGLYEAAPGYTLE 663
QY 660 DLFAATIKYVYEHPRPHSNDGKNASHEVYLGKDHSDPKNFKADEE-----PVEET 713
Db 664 DLATVYKYYVYEHPRPHSDNGFGNASDHYORNNKGADTNQTEKPESEKPOTEKPEET 723
QY 714 -----PAPEPVQVETEKVPAOLKEAEVLLAKYTDSLKANA 750
Db 724 PREKPESEKPESEKPESEPESEPESEPESEPESEPESEPESEPESEPESEPESE 783

```



PS Disclosure: Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein  
CC antigens (II) from Streptococcus pneumoniae. The protein antigens  
CC have bactericidal activity. The nucleic acids, encoding the protein  
CC antigens, may be used for the recombinant production of the proteins  
CC they encode. The protein antigens may then be used as vaccines for the  
CC prevention and treatment of Streptococcal infections in mammals  
CC (especially humans) which result in, e.g. meningitis, otitis media,  
CC bacteraemia and/or pneumonia. The present sequence represents a  
CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.  
XX

Sequence 820 AA:

Query Match 66.6%; Score 2772; DB 21; Length 820;

Best Local Similarity 65.0%; Pred. No.3e-196; Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNVSYYIDGKQATQKTENLPPDEVSKREGINAEQIYIKITDQGYV 59  
DB 2 SYELGRHQAGQVKKSNVSYYIDGQAGOKAENLPPDEVSKREGINAEQIYIKITDQGYV 61  
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNKKLDEIVNEVKGGYVIRVDGKYYVYLKD 119  
DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNKKLDEIVNEVKGGYVIRVDGKYYVYLKD 121  
QY 120 AAHADNVRTKEEINROKQEHSGHREGTPRNDGAVALARSQGRYTTDDGYIFNADLIED 179  
DB 122 AAHADNIRTKKEIKROKQESHSHNGGS--NDQAVVAARAQGRYTTDDGYIFNADLIED 179  
QY 180 TGDAYIVPHGDHYHYIPKNELASELAAEAFLSGRGNLSRTRYRONSNTSRKNVP 239  
DB 180 TGDAYIVPHGDHYHYIPKNELASELAAEAFLSGRGNLSRTRYRONSNTSRKNVP 239  
QY 240 SVSNPGTNTNTSNNSNTNSQASQNDIDSLKQLYKPLSQRHVESDGLVDPQAQITSR 299  
DB 229 NPAQPRLENNHLYTPPYHO--NOGENISSILRELIAKPLSEHVESDGLIFDPAQITSR 287  
QY 300 TARGAIVPHGDHYHYIPYSOMSELEERTARIIPLYRSNHNWPPDSRPPSPQTPPEPSP 359  
DB 288 TARGAIVPHGNHYHPIPEQSELEKRLARIIPLYRSNHNWPPDSRPPSPQTPPEPSP 347  
QY 360 GPQAPNLK-IDSN---SSLVSQALVRYKVEGYVEEKGISRYVEAKDLPSSETVKNLESKL 415  
DB 348 SPQAPNPQAPASNPIDKLYKNAKRGDGYVEPENGVSRTIIPAKDLAETAGIDSKL 407  
QY 416 SKQESVSHFTLAKKENVAPRDQEFYDKAYNLLTEAHKALFYNKGNSDFQALDKLLERLN 475  
DB 408 AKQESLSHKLGAKKTDLPSSDREFYNAKAYDILLRIHODLNNKGRQVDFEALDNLRLK 467  
QY 476 DESTNKEKLVDDLAFLAPITHPELGRPNQIETDEEVRIAQDLADKYTTSDGYIFDEH 535  
DB 468 DVPSDKVLYVDLILAFAPIRHPERLGRPNQIYTTDDEIOVAKLAGKRYTDEGDYIFDPR 527  
QY 536 DIISDEGDAYTTPHMGSHWIGKDSLSDKEKYYAAQAYTKKGIILPPSPADAVKANPTGDS 595  
DB 528 DITSDEGDAYTTPHMTGSHWIKKDSLSEARAQAAYAKEGILPPSTDHODSGNTEAKG 587  
QY 596 AAATYNNRYKGEKRIPLVRLPYMVEHTEVKNKGNLIIPKDHYNHIKFAFPDQHTYKAPNG 655  
DB 588 AEATYNNRYKAKKPYLDPMAPNLQYTVENKNGSLIIPHYDHYHNKIFEMFDEGLYEAPKG 647  
QY 656 YTLDELPAFTIKYVVEHDERPHSDNGMNASEHYLG-----KDHS----- 697  
DB 648 YTLDELPAFTIKYVVEHDERPHSDNGMNASEHYLG-----KDHS----- 697  
QY 698 -----DNNKFNKADEPEVEETPAPEVPOVETEKEVQAOLKEAEVYL 738  
DB 708 ESDEKENHAGLNPSADNLYKRPSTDEETEERAEADTTDEAETPOVENSYINAKIADAERLL 767  
QY 739 AKVYDSSILKANATEETLAGLRNNLTQIMDNNSIMAEAEKLLALLKGSNPSV 790  
DB 768 EKVTDPISIRQANAMETLTGLKSSLLGTAKDNNTISAQVDSLALLKESQAPAI 819

Search completed: November 12, 2002, 14:49:03  
Job time : 72 secs

**THIS PAGE BLANK (USPTO)**



```
|||||
Db 367 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVPAKDLPSBTVKLNESKLSKÖES 426
Qy 421 VSHITTLAKKENVAPROEEDYKAYNLTTEAHKALFKNGKNSDFQALDKLERLNDESTN 480
Db 427 VSHITTLAKKENVAPROEEDYKAYNLTTEAHKALFKNGKNSDFQALDKLERLNDESTN 486
Qy 481 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIAQADKYTTSDGYIFDEHDIISD 540
Db 487 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIAQADKYTTSDGYIFDEHDIISD 546
Qy 541 EGDAYTTPMHGSHWIGKDSLSDEKVAQAAYTKEGILPPSPDADYKANPTGDSAAIY 600
Db 547 EGDAYTTPMHGSHWIGKDSLSDEKVAQAAYTKEGILPPSPDADYKANPTGDSAAIY 606
Qy 601 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDIYHNKFAWFDHTYKAPNGYTLSD 660
Db 607 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDIYHNKFAWFDHTYKAPNGYTLSD 666
Qy 661 LEATIKYVEHDEPDRSHNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 720
Db 667 LEATIKYVEHDEPDRSHNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 726
Qy 721 QVETEKVEAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTQIMDNNSIAAEKLLA 780
Db 727 QVETEKVEAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTQIMDNNSIAAEKLLA 786
Qy 781 LKGSNPSSVSKEKIN 796
Db 787 LKGSNPSSVSKEKIN 802
```

## RESULT 2

```
E98004
hypothetical protein phla [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: E98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; H
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; N
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: E98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-828 <KUP>
A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:g15458683; GSPDB:GN00174
C:Genetics:
A:Gene: phla
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 99.9%; Score 4159; DB 2; Length 828;
Best Local Similarity 99.7%; Pred. No. 3.2e-219;
Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 SYELGLYQARTYKKNRVSIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDOGYVT 60
Db 33 SYELGLYQARTYKKNRVSIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDOGYVT 92
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGGVYIKVGGYVYLKDA 120
Db 93 SHGDHYHYNGKVPYDAIIFSELLMKDPNYKLKDEDIVNEVGGVYIKVGGYVYLKDA 152
Qy 121 AAADNVRTKEELNRKQKSHQSHREGTTPNDGAVALLANSQGRYTTDDGYIFNADSIIEDT 180
Db 153 AAADNVRTKEELNRKQKSHQSHREGTTPNDGAVALLANSQGRYTTDDGYIFNADSIIEDT 212
Qy 181 GDAYIVPHGDHHTYIPKNELSASELAAEAFLISGRGNSNSTYTRRONSNDNTSRNWPVS 240
Db 213 GDAYIVPHGDHHTYIPKNELSASELAAEAFLISGRGNSNSTYTRRONSNDNTSRNWPVS 272
```

```
Qy 241 VSNPGTTNTNSNTNSOASOSNDISLKLQYKPLPSORHVESDGLVPDAQITSRF 300
Db 273 VSNPGTTNTNSNTNSOASOSNDISLKLQYKPLPSORHVESDGLVPDAQITSRF 332
Qy 301 ARGVAAPHGDHHTYIPYSOMSELEERIAIIPLRYSNMWVPSRPEQSPQPTPEPSG 360
Db 333 ARGVAAPHGDHHTYIPYSOMSELEERIAIIPLRYSNMWVPSRPEQSPQPTPEPSG 392
Qy 361 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVPAKDLPSBTVKLNESKLSKÖES 420
Db 393 POPAPNLKIDSNSLSVQLVRKVGEGYVEEKGISRYVPAKDLPSBTVKLNESKLSKÖES 452
Qy 421 VSHITTLAKKENVAPROEEDYKAYNLTTEAHKALFKNGKNSDFQALDKLERLNDESTN 480
Db 453 VSHITTLAKKENVAPROEEDYKAYNLTTEAHKALFKNGKNSDFQALDKLERLNDESTN 512
Qy 481 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIAQADKYTTSDGYIFDEHDIISD 540
Db 513 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVRIAQADKYTTSDGYIFDEHDIISD 572
Qy 541 EGDAYTTPMHGSHWIGKDSLSDEKVAQAAYTKEGILPPSPDADYKANPTGDSAAIY 600
Db 573 EGDAYTTPMHGSHWIGKDSLSDEKVAQAAYTKEGILPPSPDADYKANPTGDSAAIY 632
Qy 601 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDIYHNKFAWFDHTYKAPNGYTLSD 660
Db 633 NRYKGEKRIPLVRLPYMVEHTVEVKNGLIIPKHDIYHNKFAWFDHTYKAPNGYTLSD 692
Qy 661 LEATIKYVEHDEPDRSHNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 720
Db 693 LEATIKYVEHDEPDRSHNDGNGNASEHVLGKKHSEDPNKNFKADEPVEETPAPEVP 752
Qy 721 QVETEKVEAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTQIMDNNSIAAEKLLA 780.
Db 753 QVETEKVEAQLKEAEVLLAKYTDSSLKANATETLAGLRNNLTQIMDNNSIAAEKLLA 812
Qy 781 LKGSNPSSVSKEKIN 796
Db 813 LKGSNPSSVSKEKIN 828
```

## RESULT 3

```
D98004
histidine motif-containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D98004
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.
e, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D98004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUP>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:g15458682; GSPDB:GN00174
C:Genetics:
A:Gene: phpa
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match 77.4%; Score 3224; DB 2; Length 855;
Best Local Similarity 75.0%; Pred. No. 3.4e-168;
Matches 617; Conservative 69; Mismatches 99; Indels 38; Gaps 4;
```

```
Qy 1 SYELGLYQARTYKKNRVSIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDOGYVT 60
Db 32 SYELGLYQARTYKKNRVSIDGKATQKTEMLTPDEVSKREGINAEOIVIKITDOGYVT 91
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGGVYIKVGGYVYLKDA 120
```

```

Db          92 SHGDHYYHNGVPPDAIIISELLMKDPNYQLKDEDIISLIGGYIVKADGYYVYLKDA 151
Oy          121 AAADNVRKEELINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 180
Db          152 AAADNVRKEELINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 211
Oy          181 GAYIYVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRONSNTSRINWVS 240
Db          212 GAYIYVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRONSNTSRINWVS 271
Oy          241 VSNPGTTNTNNTSNNTSQAOSNDISLLKOLYKPLPSQRHVESDGLVFPDPAQITST 300
Db          272 VSNPGTTNTNNTSNNTSQAOSNDISLLKOLYKPLPSQRHVESDGLVFPDPAQITST 331
Oy          301 AAGVAVPHGDHYHYPYQMSLELERIARIPLRYRSNHWVDSRPEQSPQPTPEPSPG 360
Db          332 AAGVAVPHGDHYHYPYQMSLELERIARIPLRYRSNHWVDSRPEQSPQPTPEPSPG 391
Oy          361 POPAPNLK-IDSN---SSLVSQLVKRVGEGYFEEKGISRYVFAKDLPSETVKNLESKL 416
Db          392 POPAPNLKIDSNIDSKLVKAEVAVKGVGDFYFEEGNGVRYIPAKDLSETAAGIDSKLA 451
Oy          417 KOESVSHLTAKKENAVAPRDOEFYDKATYLLTEAKKALFXNKGKNSDQALDKLERLND 476
Db          452 KOESVSHLTAKKENAVAPRDOEFYDKATYLLTEAKKALFXNKGKNSDQALDKLERLND 511
Oy          477 ESTNKEKLDLAFAPITPHERLGKPNQOJETEYDEVEYRIQOLADKYTTSQGYIFDEHD 536
Db          512 VSSDKVKLVLDLAFAPITPHERLGKPNQOJETEYDEVEYRIQOLADKYTTSQGYIFDEHD 571
Oy          537 ITSDGDVAVPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSA 596
Db          572 ITSDGDVAVPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSA 631
Oy          597 AATYRVGKERIPLVRLPYWVEHTVEYKNGLLIPHKDHYINIKFAMFDDHTTYAPNXY 656
Db          632 AATYRVGKERIPLVRLPYWVEHTVEYKNGLLIPHKDHYINIKFAMFDDHTTYAPNXY 691
Oy          657 TLEDFATIKYVEHPRPSNDGNASEHVLGKDHSEDPNKFADPE-----PV 710
Db          692 SLEDLATYKYYVEHPRPSNDGNASEHVLGKDHSEDPNKFADPE-----PV 751
Oy          711 EET-----PAPEVPOVETEKVEAOLKEAELLAAVT 742
Db          743 DSLKANATETLAGLNNLLTQIMDNNSIMAEKLLALIKGS 785
Oy          812 NPIISNAKETILGLKNNLLFGTDNNNTIMAEKLLALIKGS 854
Db          854 NPIISNAKETILGLKNNLLFGTDNNNTIMAEKLLALIKGS 854

RESULT 4
895136
conserved domain protein Sp1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B95136
R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap-
ple, Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:914972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: Sp1174

```

```

C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein
Query Match          67.2%; Score 2797.5; DB 2; Length 819;
Best local similarity 67.0%; Pred. No. 5.9e-145;
Matches 546; Conservative 84; Mismatches 136; Indels 47; Gaps 7;

Oy          1 STELGLYQA-RIVYKNNRVSYIDGKOATQNTENAPDEVSKREGINAEDIVIKITDGYV 59
Db          21 SYELGYQAQODKESNRVAIIDGQAKAENLTFDESVKREGINAEDIVIKITDGYV 80
Oy          60 TSHGDHYHYNKVPDAIIISELLMKDPNYQLKDEDIISLIGGYIVKADGYYVYLKDD 119
Db          81 TSHGDHYHYNKVPDAIIISELLMKDPNYQLKDEDIISLIGGYIVKADGYYVYLKDD 140
Oy          120 AAADNVRKEELINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIE 179
Db          141 AAADNVRKEELINROKQOESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIE 197
Oy          180 TGDVAVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRONSNTSRINWVP 239
Db          198 TGDVAVPHGDHYHYPKNELSASELAAPFLSGRGNLSNSTRYRONSNTSRINWVP 246
Oy          240 SVSNPGTTNTNNTSNNTSQAOSNDISLLKOLYKPLPSQRHVESDGLVFPDPAQITSR 299
Db          247 NPAPRLSENHMLTYPTYHQ--NOGENISLRLRELYAKPLSERHVESDGLVFPDPAQITSR 305
Oy          300 TARGVAVPHGDHYHYPYQMSLELERIARIPLRYRSNHWVDSRPEQSPQPTPEPSP 359
Db          306 TARGVAVPHGNHYHYPYQMSLELERIARIPLRYRSNHWVDSRPEQSPQPTPEPSP 365
Oy          360 GPQAPNLKIDSNSSLVSQLVKRVGEGYFEEKGISRYVFAKDLPSSETVKNLESLSKOE 419
Db          366 GPQAPNLKIDSNIDSKLVKAEVAVKGVGDFYFEEGNGVRYIPAKDLSETAAGIDSKLAGE 423
Oy          420 SVSHLTAKKENAVAPRDOEFYDKATYLLTEAKKALFXNKGKNSDQALDKLERLNDST 479.
Db          424 SVSHLTAKKENAVAPRDOEFYDKATYLLTEAKKALFXNKGKNSDQALDKLERLNDST 483
Oy          480 NKEKLVLDLAFAPITPHERLGKPNQOJETEYDEVEYRIQOLADKYTTSQGYIFDEHDIS 539
Db          484 DKVKLVLDLAFAPITPHERLGKPNQOJETEYDEVEYRIQOLADKYTTSQGYIFDEHDIS 543
Oy          540 DEGDAYVTPPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSAAI 599
Db          544 DEGDAYVTPPHMGSHWIGKDSLSDEKVAQAQYTEKGIPLPSPADYKANFTGDSAAI 603
Oy          600 YNRVAGKERIPLVRLPYWVEHTVEYKNGLLIPHKDHYINIKFAMFDDHTTYAPNXY 659
Db          604 YNRVAGKERIPLVRLPYWVEHTVEYKNGLLIPHKDHYINIKFAMFDDHTTYAPNXY 663
Oy          660 DLFATIKYVEHPRPSNDGNASEHVLGKDHSEDPNKFADPE-----PVET 713
Db          664 DLFATIKYVEHPRPSNDGNASEHVLGKDHSEDPNKFADPE-----PVET 723
Oy          714 -----PAPEVPOVETEKVEAOLKEAELLAAVT 750
Db          724 PREEKQSEKPEPSPKPTPEPEPSEPEQVETEKVEKLEAEDLLGKIDPILKNSA 783
Oy          751 TETLAGLNNLLTQIMDNNSIMAEKLLALIKGS 785
Db          784 KETTLGLKNNLLFGTDNNNTIMAEKLLALIKGS 818

RESULT 5
695115
conserved hypothetical protein Sp1003 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Teitelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei-
son, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzap-
ple, Science 293, 498-506, 2001

```

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.;  
 A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*.  
 A:Reference number: A55000; MUID:21357209; PMID:11463916  
 A:Accession: G95115  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-839 <KUR>  
 A:Cross-references: GB:AE005672; PIDN:AAK75120.1; PID:g14972476; GSPDB:GN00164; TIGR:SPA  
 A:Experimental source: strain TIGR4  
 C:Genetics:  
 A:Gene: SPI003  
 C:Superfamily: *Streptococcus agalactiae* hypothetical 92.4k protein

Query Match 66.6%; Score 2772; DB 2; Length 839;  
 Best Local Similarity 65.0%; Pred. No. 1.5e-143;  
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

```

QY 1 STELGLYQARTV-KENNRYSYIDGKQATQKTENLPDEVSKREGINAEOIVIKITDQGYV 59
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 21 SYELGRHOAGQVKKESNRYSIDGQAGKAENLTPDEVSKREGINAEOIVIKITDQGYV 80
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGKGYIVDQKYYVYLLD 119
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGKGYIVDQKYYVYLLD 140
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 120 AAHADVNRKKEELNRKQKSHQREGTFRNDGAVVALASQGRYTTDDGYIFNASDIIED 179
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 141 AAHADNIRKKEELRKQKSHSHHGSGS--NQDAVVAARAQGRYTTDDGYIFNASDIIED 198
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 180 TGDVAVPRGHDHYTPKMLASSELAAEAFLSGRGNLSNRTTYRQNSDNTSRTNWVP 239
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 199 TGDVAVPRGHDHYTPKMLASSELAAEAFLSGRGNLSNRTTYRQNSDNTSRTNWVP 247
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 240 SVSNPQTNTNTSNNTSNQASQNSDIDSLKQLYKLPLSORHVESDGLVDPQAITSR 299
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 248 NPAQRLSENHNLTVPTVHQ--NQGEMISLLRELTAKLSEHNVSDGLIEDPQAITSR 306
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 300 TARGVAVPRGHDHYTPYQMSSELEERIRIIPLRKRSNHWVPDSRPEQSPQPTPEPSP 359
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 307 TARGVAVPRGHDHYTPYQMSSELEERIRIIPLRKRSNHWVPDSRPEQSPQPTPEPSP 366
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 360 GQOPARNLK-IDSN---SSLVSQLVKRVGEGYVEEKGISRYFAADLSEYTKNESKL 415
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 367 SPQAPNPQAPSPNPIDEKLVKVAARKVGDGYVEENGVPRTIIPAKDLAEFTAAGIDSKL 426
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 416 SKOESVSHLTAKKENVAPRDQEFYDKAVNLTLEAKALFXNKGSRNSDFOALDKLERLN 475
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 427 AKQESISHLGAKKTDLPSSDRFYKAYDLARIHQDLIDNKGROYDEALDNLLERLK 486
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 476 DESTNKEKLVDDLAFADITPBERLGKPNQSEYTEDEVRIQAOLADKYTSDGYTFDEH 535
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 487 DVPSDKVKLVDDLAFADIRPBERLGKPNQAITYTDDEIQVAKLAGKYTEDEGYTFEPR 546
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 536 DIISDGDVAVPRPHMGSHWIKGDSLSDKEKYAAQAAYTEKGLIPSPSPADVAKANPTGOS 595
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 547 DIISDGDVAVPRPHMGSHWIKGDSLSEAEARAQAAYAKKEGTLTPSPSTHODSGNTEAKG 606
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 596 AAIYRVYRGEKRIPLVRLPYWEHTEVEKNGNLIIPHODHYNIKFAFDDHTYAPRG 655
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 607 AAIYRVYRGAARKVPLDRMPYNLQYVEVKNGLIIPHODHYNIKFEWDESLYEPARG 666
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 656 YTLLEDLFAITIKYVEHPDRPHSDNGMNASEHYLGK-----KDHSF----- 697
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 667 YTLLEDLFAITIKYVEHPDRPHSDNGMNASEHYLGK-----KDHSF----- 726
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 698 -----DPNKNFADEEVEETPAEPEVPOVETEKYEAOLKEAEVLL 738
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 722 ESDKENHAGLNPSADNLKPSPTDETEBEAEDTDEAEIPQEVSVINAKIADAELL 786
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 739 AAVTOSSTLANATELTLAGLRNNLTQIMDNNSIMAEKLLMLKSGNSSSV 790
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 787 EKVYTPSIRONAMEYTLTGKSSLLGTCKDNNTISAEDVSLALLKESQAPPI 838
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  
```

RESULT 6  
 C97985  
 hypothetical protein phd [imported] - *Streptococcus pneumoniae* (strain R6)  
 C:Species: *Streptococcus pneumoniae*  
 C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 02-Nov-2001  
 C:Accession: C97985  
 R:Host(s): J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.  
 e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.  
 Y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
 A:Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.  
 A:Reference number: A97872; MUID:21429245; PMID:11544234  
 A:Accession: C97985  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-853 <KUR>  
 A:Cross-references: GB:AE007317; PIDN:AAK99711.1; PID:g15458514; GSPDB:GN00174  
 C:Genetics:  
 A:Gene: phd  
 C:Superfamily: *Streptococcus agalactiae* hypothetical 92.4k protein

Query Match 65.6%; Score 2733; DB 2; Length 853;  
 Best Local Similarity 63.5%; Pred. No. 2.1e-141;  
 Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;

```

QY 1 STELGLYQARTV-KENNRYSYIDGKQATQKTENLPDEVSKREGINAEOIVIKITDQGYV 59
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 21 SYELGRHOAGQVKKESNRYSIDGQAGKAENLTPDEVSKREGINAEOIVIKITDQGYV 80
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGKGYIVDQKYYVYLLD 119
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 81 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGKGYIVDQKYYVYLLD 140
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 120 AAHADVNRKKEELNRKQKSHQREGTFRNDGAVVALASQGRYTTDDGYIFNASDIIED 179
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 141 AAHADNIRKKEELRKQKSHSHHNS--RADAAVAAARAQGRYTTDDGYIFNASDIIED 197
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 180 TGDVAVPRGHDHYTPKMLASSELAAEAFLSGRGNLSNRTTYRQNSDNTSRTNWVP 239
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 198 TGDVAVPRGHDHYTPKMLASSELAAEAFLSGRGNLSNRTTYRQNSDNTSRTNWVP 246
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 240 SVSNPQTNTNTSNNTSNQASQNSDIDSLKQLYKLPLSORHVESDGLVDPQAITSR 299
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 247 NPAQRLSENHNLTVPTVHQ--NQGEMISLLRELTAKLSEHNVSDGLIEDPQAITSR 305
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 300 TARGVAVPRGHDHYTPYQMSSELEERIRIIPLRKRSNHWVPDSRPEQSPQPTPEPSP 359
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 306 TARGVAVPRGHDHYTPYQMSSELEERIRIIPLRKRSNHWVPDSRPEQSPQPTPEPSP 365
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 360 GQOPARNLK-IDSN---SSLVSQLVKRVGEGYVEEKGISRYFAADLSEYTKNESKL 415
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 366 SPQAPNPQAPSPNPIDEKLVKVAARKVGDGYVEENGVPRTIIPAKDLAEFTAAGIDSKL 425
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 416 SKOESVSHLTAKKENVAPRDQEFYDKAVNLTLEAKALFXNKGSRNSDFOALDKLERLN 475
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 426 AKQESISHLGAKKTDLPSSDRFYKAYDLARIHQDLIDNKGROYDEALDNLLERLK 485
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 476 DESTNKEKLVDDLAFADITPBERLGKPNQSEYTEDEVRIQAOLADKYTSDGYTFDEH 535
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 486 DVPSDKVKLVDDLAFADIRPBERLGKPNQAITYTDDEIQVAKLAGKYTEDEGYTFEPR 545
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 536 DIISDGDVAVPRPHMGSHWIKGDSLSDKEKYAAQAAYTEKGLIPSPSPADVAKANPTGOS 595
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 546 DIISDGDVAVPRPHMGSHWIKGDSLSEAEARAQAAYAKKEGTLTPSPSTHODSGNTEAKG 605
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 596 AAIYRVYRGEKRIPLVRLPYWEHTEVEKNGNLIIPHODHYNIKFAFDDHTYAPRG 655
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 606 AAIYRVYRGAARKVPLDRMPYNLQYVEVKNGLIIPHODHYNIKFEWDESLYEPARG 665
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 656 YTLLEDLFAITIKYVEHPDRPHSDNGMNASEHYLGK-----DHSDEPNK----- 701
    |||||:..|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
  
```



Db 666 YSLEDLATVYVYHPNERNRPHSDNGFCMASDHVRKNGQADTNQTEKPNNEKQTEKP 725  
Qy 702 -----NEKAD-----EPPVEPFAPEVPOVE 723  
Db 726 EEDKEHDEVSEPTPESEKENHNGLNPSADNLTKPSTDTETEEDADDTDEAEITPOVE 785  
Qy 724 TEKEVAOLKEAEVLLAKVYDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLALLK 783  
Db 786 HSVIAKIAEAEALLEKTYDSSIRQNAVETLGLKSSLLTGKONNTISAEVDSLALLK 845  
Qy 784 GSNBSSV 790  
Db 846 ESQPTPI 852

## RESULT 7

conserved hypothetical protein SP1004 [imported] - Streptococcus pneumoniae (strain TIGR H95115  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 03-Aug-2001  
C:Accession: H95115  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidon, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, nson, T.; Hickey, E.R.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A.; Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: H95115  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1039 <KUR>  
A:Cross-references: GB:AE005672; PIDN:AAK75121.1; PID:g14972477; GSPDB:GN00164; TIGR:SP4  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP1004

Query Match 29.9%; Score 1246; DB 2; Length 1039;  
Best Local Similarity 40.9%; Pred. No. 3.6e-60;  
Matches 296; Conservative 97; Mismatches 180; Indels 150; Gaps 18;

Qy 1 SYELGLQARTVKNRNVSYDGRKQATQKTEMLPDEVSKREGINAEOIVIKITDQGYT 60  
Db 22 AYALNQRHSQENKNNRVSVDGSGSQSENLPDQVSKREGIOAEQIVIKITDQGYT 81  
Qy 61 SHGHYHYNGKVPYDALISELLMKDPNYKTKDEDIVNEVKGGYVTKVDGKYVYVYLKDA 120  
Db 82 SHGHYHYNGKVPYDALISELLMKDPNYKTKDEDIVNEVKGGYVTKVDGKYVYVYLKDA 141  
Qy 121 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 180  
Db 142 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 197  
Qy 181 GDATVPRGHGHHYIPKNELSASELAAAEAFLSGRGLNSRRTYRNRONSNTSRTNWPS 240  
Db 198 GNAVTPRGHHYIPKSDLSASELAAGAKHAKNMQPSOLSYSTFASDN----- 248  
Qy 241 VSNGTGTNTNNSNNSQASQSDNDISLLKOLYKPLSQRHESQGLVDPDAOITSRT 300  
Db 249 -----NTQSVAKGSTSKPANKSENLQSLKELTDSPSAORYSESDGLVDPDAKIIISRT 301  
Qy 301 ARGVAVPHGHHYIPYSOMSELEERIALIPLRYRSNHWPPDSRPEQSPQPPPEPSPG 360  
Db 302 PNGVAIPHGHHYIPYSKLSALEEKIARMP----- 333  
Qy 361 POPAPNLKIDNSSLSVLQVAKVGEVFEKGISRYVFAKOLPSEFYVKNLESKLSQES 420  
Db 334 -----ISGTGSTVSNNAK-----PNEVYVSSIGLSLSPSS 363  
Qy 421 VSHTLTAKKEVAVRQDFYKAVNLTLEAHKALEFXNKGGRNSDQADLKLRLINDESTN 480  
Db 364 -----LITSKELSSASDGIITNPK-DIVEETATATYVRHG--DHFHYTPK-----SNOIG 410

Qy 481 KEKIVDOLLAFLAPITPHERLGLKPNQSIETYEDEVARIAQLADKYTSDGYIFDEHDIISD 540  
Db 411 QPTLPNNSLATPSP-SLPIINGTSHKEHE-----EGCYGDFARRIIAE 452  
Qy 541 EGDAYVPRHGMHSHWIGKDSLSDEKEVAAQAYTEKGIPLPSPDADVAKANPTGDSAAAY 600  
Db 453 DESGVMSHGHNNHYFFKDKLTDEQIKAAOKHLEE-----VYTSNNGDLDSLSSH 501  
Qy 601 NR-----VKEKRIPVLRPLPYW--EHTVEVKNGLIIT-PKHCHYHNKIRAM 644  
Db 502 EDDYPSNAKEKMDLQKTEEKIAGIMQYGVKRSIVNKKERMAITYPHGDHHAADP--- 558  
Qy 645 FDDHTYKAPNGYTLEDDPATIKYVYHPDERPHSNDG-----WGNASEHYLKKDKHSEDPN 700  
Db 559 IDEH---KPVGIG---HSHSNYELFKPEBGVAKKEGKKYVTGBELTNVYVNLKRNSTFNN 611  
Qy 701 KNF 703  
Db 612 QNF 614

## RESULT 8

hypothetical protein phec [imported] - Streptococcus pneumoniae (strain R6)  
C:Species: Streptococcus pneumoniae  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 22-Oct-2001  
C:Accession: D97985  
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgelt, S.; DeHoff, B.S.; e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.  
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A:Reference number: A97872; MUID:21429245; PMID:11544234  
A:Accession: D97985  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1039 <KUR>  
A:Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:g15458515; GSPDB:GN00174  
C:Genetics:  
A:Gene: phec

Query Match 29.8%; Score 1243; DB 2; Length 1039;  
Best Local Similarity 41.2%; Pred. No. 5.3e-60;  
Matches 298; Conservative 94; Mismatches 180; Indels 152; Gaps 20;

Qy 1 SYELGLQARTVKNRNVSYDGRKQATQKTEMLPDEVSKREGINAEOIVIKITDQGYT 60  
Db 22 AYALNQRHSQENKNNRVSVDGSGSQSENLPDQVSKREGIOAEQIVIKITDQGYT 81  
Qy 61 SHGHYHYNGKVPYDALISELLMKDPNYKTKDEDIVNEVKGGYVTKVDGKYVYVYLKDA 120  
Db 82 SHGHYHYNGKVPYDALISELLMKDPNYKTKDEDIVNEVKGGYVTKVDGKYVYVYLKDA 141  
Qy 121 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 180  
Db 142 AHADNVRTKEINRQKQESHQREGGTPRNDGAAVALARSQGRYTTDDGYIFNADIIEDT 197  
Qy 181 GDATVPRGHGHHYIPKNELSASELAAAEAFLSGRGLNSRRTYRNRONSNTSRTNWPS 240  
Db 198 GNAVTPRGHHYIPKSDLSASELAAGAKHAKNMQPSOLSYSTFASDN----- 248  
Qy 241 VSNGTGTNTNNSNNSQASQSDNDISLLKOLYKPLSQRHESQGLVDPDAOITSRT 300  
Db 249 -----NTQSVAKGSTSKPANKSENLQSLKELTDSPSAORYSESDGLVDPDAKIIISRT 301  
Qy 301 ARGVAVPHGHHYIPYSOMSELEERIALIPLRYRSNHWPPDSRPEQSPQPPPEPSPG 360  
Db 302 PNGVAIPHGHHYIPYSKLSALEEKIARMP----- 333  
Qy 361 POPAPNLKIDNSSLSVLQVAKVGEVFEKGISRYVFAKOLPSEFYVKNLESKLSQES 420  
Db 334 -----ISGTGSTVSNNAK-----PNEVYVSSIGLSLSPSS 363

[illegible][illegible]



```

Db 784 ONSQSVVEEIPVNEDEFEEDVHTDQLDHLKTVDPVEIPEVEEIPSELHENVAPHEIETE 843
Oy 469 KLERLNDSTNKKELVDLLAFLA-PTRPERLGRPSQIETDEE-VRIQLOADKRYT 526
Db 844 EYFPEPNQNNNEQOINEDDKSAHIOHEIVEVEIILPEDKNKEVEHEIVEVEIILPEDKN 903
Oy 527 SDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDEKREVAQAAYTKRGILPPSPDAD 586
Db 904 EKG-----QHEIVEVE---ELPE-----DDKNEKEVEHEIVEVEIILP--EDKN 942
Oy 587 VKANPTGDSAAAIYNRVKEGKRIPLVRLPYMEH-TVEYKKNGLIIPHDHYNHKKFAMF 645
Db 943 EKGQHEIVEVEIILPEDKNKE-----VEHEIVEVEE---ILPEDKNKG----- 983
Oy 646 DHTYKAPNGTYLEDLFATIKYVEHPDE-RPHSNDGMCNAS-----EHVLGKDHSE--- 697
Db 984 -QHEIVEVEIILPEDKNKEVQHEIVEVEIILPEDKNKEQHEIVEVEIILPEDKNKEQ 1042
Oy 698 -----DPKNKFADEPEVEETPAPEVPOVETKEVEAQLKEAVLLAKVYDSSLK 747
Db 1043 HEIVEVEIILPEDKNKEVQHEIVEVEIILPEDKNKEVQHEIVEVEIILPEIIVE---- 1095
Oy 748 AATETLACLRNNLTQIMDNNSIAAEKLLALLKGSNPSGVSKKI 795
Db 1096 --TEEVPSTNNNENI-----TIKEPK-----NEFSVEEKAI 1128

```

# RESULT 13

Fnth protein [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
 C:Accession: D90011  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
 ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hirataetsu, K.  
 Lancet 357, 1225-1240, 2001  
 A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: D90011  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-2481 <KUR>  
 A:Cross-references: GB:BA000018; PID:g13701961; PIDN:BA043253.1; GSPDB:GN00149  
 A:Experimental source: strain N315  
 C:Genetics:

A:Gene: fnbB(mrp)  
 Query Match 4.1%; Score 172; DB 2; Length 2481;  
 Best Local Similarity 19.4%; Pred. No. 0.5;  
 Matches 169; Conservative 140; Mismatches 353; Indels 210; Gaps 38;

```

Oy 8 OARTYKENVRSYIDGKQATOKT-----ENLPPDEVSKREGINMAQIYIKITDGYVTS 61
Db 969 QAATTTKSAPKAEIAOKASEKRTALEAMNDSTTEEOAKKDV-----DQAVVTA 1018
Oy 62 HSD-----HYHYNGKVPYDAIISEELLMKDPNFKLKDEDIVNEVKGGYIVKDGKYYV 115
Db 1019 NADIDWATANTOVDAKKTNEATIA--AITPDANVPRAKQAIAD-----KVQAO--- 1066
Oy 116 YLKDAHADNVRTKEINRQKQDESHRGSGTRPRNDGAVALAR-----159
Db 1067 --ETIDANGSTTEKEAKAOVOTEKTA---DAAIDAASHNVEVEAKNAETAKIE 1120
Oy 160 --QGRYTDGTYFNASDIIEDTGDAIYIPHDHVIYIKNELSASELAAAEFLSGRN 217
Db 1121 AIQPATTKD---NAKQAIATKANERKTA-----IAOTQITAEELIAANA-----D 1164
Oy 218 LSNRTYRQNSDNTSRTVMVPSVSNPTGTTNTNTNSNTNSQASQSDIDSLKQLYKL 277
Db 1165 VNATYQANSNIENANSQNDVQAKTGETSTD-QVTPYVNNKATARNITAILNN--KL 1221
Oy 278 PLQSRHVESDGLVFDPQITSRARGAVVPHGDHYHIFIPYSQMSLEERIAIIPLRYS 337

```

```

Db 1222 QRIQATPPATDEKQAAAEANTENGA-----NOAISAAITNAQVDEAKANA 1269
Oy 338 NIMVPSDREQSPQPTPEPPSPGQPA-----PNLKIDNSSLVQYRKVEG--- 386
Db 1270 EAINAVTPKVVYKQAADEIDQLOATOTVINNDQNMATNEKEKAIDQALATVAIDAKNN 1329
Oy 387 --YVEEKGISRYFAKDLPESETVNLSS---KLSQESVSHTLAKRENAVAPROQEFY 440
Db 1330 ITAATDDKGVDT---AKQAGKNSIOSTQPATAVKSNAKNEVQAVTQNOAL----- 1378
Oy 441 DKAYMLTEPAHKALYXNGKNSDFQALDKLER-LNDSTN-----KEKYDDLAFIAP 494
Db 1379 DMTTGATTEE-----KNAKDLVLAKREKAYODIINAOITNDVTOIKQOAAVDIGITAD 1433
Oy 495 ITHPERLGKPSQIETDEEVRIQLOADKRYTSDGYIFDEHDIISDEGDAYVTPHMGSH 554
Db 1434 TTIKD-VAKDELATRANEKALIAOTADAT-----EKEQANOQVADQLT--QGNON 1483
Oy 555 WIGKDSLDEKREVAQAAYTKRGILPPSPDADYKANPTGDSAAAIYNRVKEGKRIPLVRL 614
Db 1484 IENASIDVNTAKDNAL---QALDPIQASTFDVKTNARAEILTEMQNKI-----TEI 1532
Oy 615 PYMVEHYEVKKNGLIIP-----HKDHYHKKFAMF--DHTYKAPNGTYLEDLFATIKY 668
Db 1533 LNNNETTEEEK-GNDIGPRAAYEGGLNNINATTTGQVTTAKDPAVOKVQDLHA----- 1586
Oy 669 VEHPDERPHSNDGMCNASEHVLGKKDHSEDPKNKFADEE-PVEETP--AEPEVPOVETE 725
Db 1587 --NPYKKP-----AGKKELDQ-----AAADKKTQIEQTPNASQOEINDADQOE 1626
Oy 726 KYEAOULKEA-----EVLLAKVYDSSLKANAETETLAGLRNNLTQIAMD----- 767
Db 1627 -VDTEINQAKTAVNDOSSTNEYVDNAVKEGKAKINAVKTFSEYKKDALAKIEDAVNAKYNE 1685
Oy 768 -----NNSIAEAEKLLALLKGSNPSVSK 792
Db 1686 ADNSNASTSETAEAKQIAELKQTADQNVNQ 1717

```

# RESULT 14

hypothetical protein YF016c - yeast (Saccharomyces cerevisiae)  
 C:Species: Saccharomyces cerevisiae  
 C:Date: 02-Sep-1995 #sequence\_revision 19-Oct-1995 #text\_change 19-Apr-2002  
 C:Accession: S56271  
 R:Murakami, Y.; Naitou, M.; Hagihara, H.; Shibata, T.; Ozawa, M.; Sasamura, S.I.; Sas  
 submitted to the EMBL Data Library, May 1995  
 A:Description: Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces*  
 A:Reference number: S56186  
 A:Accession: S56271  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1233 <MUR>  
 A:Cross-references: EMBL:D50617; NID:g836685; PIDN:BA09255.1; PID:d1009896; PID:g836  
 C:Genetics:  
 A:Map position: 6R

Query Match 4.1%; Score 171; DB 2; Length 1233;  
 Best Local Similarity 19.8%; Pred. No. 0.21;  
 Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;

```

Oy 12 VKENNRVSYIDGKQATOKTENTLPDEVSKREGINAEQIYI-----KITD-QGYVTS 61
Db 151 IKETSTNNVAEG-----TENVP--IKESTGIEVGNSPITPRKKKKKKTTNRGRSS 202
Oy 62 HGDHYHYNGKVPYDAIIT--SEELLMKDPNFKLKDEDI-----VNEVKGGYIVVD 110
Db 203 NPADTDLSTKOSTLDSILVIGIEVLEQDGS---KNEDIKVNIYQDEQVAVNEKMDIRTRNE 259
Oy 111 GKIYYLLKDAHADNV-----RTKEINRQKQDESHRGSG--TPRNDGAVALARSQ--- 160

```

```

Db 260 SSDKTFIDIDVKNKNDVETSSKSENNINEEKAHTLPPRENEILNVNGNAASEFKHOLEP 319
QY 161 -GRYTTDDGYIFNNSADLIEDTGDAYIVPHGDHYHYIPKNETLSASELAAPLSCGNMIS 219
Db 320 HGLEAGDE-----NGASATKDYSESLTNGFNFKENESKHLKAGKQOTE---SPRDIS 372
QY 220 NSRTRYRONSNTSTNTNVPVSNP-----GTTNTNTSNSTNSQASQSNDI----- 267
Db 373 PSVLAKNOKETEIGEDHVFQKDEKCRKELSVNHNHNFNAAGSDSIIPETE 432
QY 268 -----DILLKQLYKLPISORH-----ESGLVFDPAQ---ITSR 299
Db 433 RETYDETMGPTKRISDEKMLQHGNDISVEEKEEEEEENSTFSYKKNENVTGE 492
QY 300 -----TARGVAVPHGDHYHFLPYSONSELEPRARIIPLRYSN----- 338
Db 493 QEAVRNNNEVSGTEESTSKGEIIMGD-----EKQSEAGPK-SSTIEIGSANSATIS 544
QY 339 --HWVPDSRPQSPQPPPEPSPQAPNLKIDSNSLSVQLYRKVGEYVFEKGISIR 396
Db 545 KDNVLDEDEAPQENKPTVEYGEIDIPDAPRD-----VEIYAVAKNLIIPEDLEVAK 599
QY 397 -----YVAKDLPSSTVYNLESKLSKQSVSHTLTAKENAVP-----DOEYDK 442
Db 600 EDQEGEYQKLPDEPVKAMDKIAMRGASISIDMKKKQEGTAEISNEKAKKEVEDTARES 659
QY 443 AYNLTEAHKALFYNK-----GRNSDFOALDKLERLNDSTKKEKLYVDLLAFIAP 495
Db 660 AEGVVEKSKTPESPESKRYKCTSGRPEDLOJNERPEVLKEDVRYPDDVAPETATTEN 719
QY 496 THPERLGNPSQIETEDEVRIADLADKYTTSQYIFDEHD-----TISDEGDAYVTPHM 550
Db 720 SEEDPKSQROVQISTEQAEITQKMDGVGFTS---FKEEKPKRFETIQGDKITGKDT 776
QY 551 GHSHWIGDSISDEKVAQAQAYTEKKGILPS-----PDADYKANPTODSAAAIYNR 602
Db 777 NHEGGEATEAASENSK-ASDVGTAKYITEPSSSEYKKDTEBDAEAE-----NSEKTEIK 830
QY 603 VKGEKRIPLVAPVVEHTEVKNKGNLTI-----PHKDHYNIKFAMFDHDTY 650
Db 831 VKAE--LENDAPKPAEYTALENKENEDEVTEDEDAVEENSEKTEFKYK--AELGNL 885
QY 651 KAPN--GYTLE-----DLFATIKYVE---HPDERPHSNDGWNASBHVLAGKD-- 694
Db 886 DAPKAEAYTALENKENEDEVTAASKEDIETKSPAPTEIED--GCTEABVSKKNAE 942
QY 695 --HSDPKKNKKADEPEVEETPAPEVQVE-----TEKVE-AOLKEAVLLAVYTSSL 746
Db 943 AVTKEDENMENSIAEALKDVTGDOEDIDINISDEFORTVELPELEKODIDKNKGDEKL 1002
QY 747 KANATELAGL-----RNLTFLQINDNNSIMAAEKLALALGKSPSSVSKKIN 796
Db 1003 EVELTEKETSIPDLVEENIT---EKKNEIKOEBE-VSOLDENETESISKAPN 1053

RESULT 15
D84900
hypothetical protein At2g46240 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: D84900
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanacker, S.E.; Umayam, L.; Tallon, L.;
Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402:761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: D84900
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1043 <STO>
A:Cross-references: GB:AE002093; MID:93702325; PIDN:AA062882.1; GSPDB:GN00139
C:Genetics:

```

A:Gene: At2g46240  
A:Map position: 2

Query Match 4.08; Score 168.5; DB 2; Length 1043;  
Best Local Similarity 18.58; Pred. No. 0.22;  
Matches 169; Conservative 108; Mismatches 327; Indels 311; Gaps 38;

```

QY 40 KREGINAOIYKTIPOQGYTSHGDHYHYNGV---PYDAIISBELMKDPNLYKDED 96
Db 271 KRQVYASE-----SSNEDRKKMGKQKIYEFYFIDISMIKSL----- 306.
QY 97 IVNEKGGYVIAKVDGKYVYLKDAHADNVRKEDI-----NRQKQHSOH 142
Db 307 ---IQG-----QDYKEAQONQKNKEEPQGVPIPIFWIPSYGRKKDVEASES 348
QY 143 REGGTP-----RNDCAVALARSQGRYTTDGYIFNNSADLIEDTGDAYIVPHGD 190
Db 349 KESSNEGHNLESCPSDLHRNEGQITQAKGK-----EGNFECNVLSDAEKKSVINIPIVAN 403
QY 191 HY-----HYIPKNETLSASELAAPLISGR----- 215
Db 404 HLOEPRNIPVKLSENHLEKPTPEPTKRIAKNEPVSKTKQSSSSSEASKLPPVCLIRDP 463
QY 216 -----GNLSNSRTYRONSNDNT-----SRTNWPVSYNPGTNTNTS---NN 254
Db 464 PKERNGSKSVSHPRKMKESKETKIAAPLSSKKAESRT--VPEACNVKCEDANMEMKMAE 521
QY 255 SNTNSQASQSDIDSLKQLYKLPISORHVESDGLVFPDPAQITSTRTAGVAVPHGDHYH 314
Db 522 GSLNALRTEKGSVES-----NSNLQSESNGEIIRPKCAKENREOP----- 561.
QY 315 IYPSOMSELEPRARIIPLRYS---NHVPDSRPQSPQPPTEPPSPG----- 360
Db 562 ---AKSTTEEARIIISMTRGYDVRKRPDIKKIETAI---TYREQGDYKKRIEALFA 615
QY 361 --POPAPNLKIDSNSLSVQLYRKVGEYVFEKGISIRYVAKDLPSSTVYNLESKLS 416
Db 616 STDGHIIEKEIIVNGELVWMLLKLD---AVEGLHPSIREFRKALATE-LSSIQDKLSD 671
QY 417 KOESVSHTLTAKENAVARDFEDYDAVNL---LTEAHKALFYNKGNPSQALDKLLE 472
Db 672 KNSCASAEKEVAKQVEIKSQP-SDSPVNLHSQLTLENKAV----- 712
QY 473 RLNDSTKKEKLYVDLLAFIPIHPERLGNPSQIETEDEVRIADLADKYTTSQYIF 532
Db 713 ---SDTNLEKYLVR-----LSPEEH-----PMSVLANRTDEK---QAESAAETEEG 751
QY 533 DEHDIIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQAYT-----KEKGILPPSPDA 585
Db 752 -----GLFETLATDSKQATENAAAASSTTIEKIGEVYTVPGNP-- 791
QY 586 DYKANPTGDSAAAIYNRYKGEKRI-----PLYRLPVMEHT-----VEY-- 624
Db 792 ---PSADGNGMTVTVTEENKAMAVESLEPIINELPQMVETETNSIRDPENASVSEA 846
QY 625 -----KNGLLIIPKHD---HYNNIKFAMPDHDT---YKAPNGYTLLEDLFA-- 663
Db 847 ETNSSENNENKRGEDDIYVHSEKNVELSELPGVYIDEETQPLSQDSSSYTREGMKTAMP 906
QY 664 -TIKYVEHPDERPHSNDGWNASEHVLGKKDHSDEPNKKNKADEPEVE-----ETPAE 716
Db 907 KTAQOEETEVDSHNSNKGIGQQTSEPODEKEGS--PETEIVYEQGLETFEVLINBGAPE 964
QY 717 PEYQVETEKVEAOLKAEBVLLAVYTSSKANATETLAGIRNLTQINDNNSIMAAEAE 776
Db 965 PETHEPGISKETKILMEENQFKEETMETIVKAGREQ--LEVISKLTSRVXSLERKLSHK 1022
QY 777 KILALLKGSNPSVS 791
Db 1023 KTOIRRRASKPMVS 1037

```

Search completed: November 13, 2002, 04:11:22

Job time : 55 secs

---

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 14:49:09 ; Search time 25 Seconds

(without alignments)  
1320.606 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SEIGLYQARTVKEKNRVS.....KLALKGNPSVSKEKIN 796

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	4.1	1233	YF16_YEAST	P43597 saccharomyc
2	161	3.9	1420	APX_XENLA	Q01613 xenopus lae
3	158	3.8	1118	UBP8_HUMAN	P40818 homo sapien
4	157	3.8	910	DNJM_MYCPN	P73354 mycoplasma
5	157	3.8	1658	YMG7_YEAST	Q03661 saccharomyc
6	157	3.8	2195	SC16_YEAST	P48415 saccharomyc
7	156.5	3.8	1185	MAPX_DROME	P23226 drosophila
8	154.5	3.7	3924	ANK2_HUMAN	Q01484 homo sapien
9	154	3.7	1453	Y373_BOVIN	Q91u23 bos taurus
10	153.5	3.7	1539	Y373_HUMAN	O15078 homo sapien
11	150.5	3.6	1176	YOH8_YEAST	Q08236 saccharomyc
12	149.5	3.6	1175	PTNL_RAT	O62728 rattus norv
13	149	3.6	2869	RBL1_PLAYB	Q00798 plasmodium
14	145.5	3.5	1085	CUT7_SCHPO	P24339 schizosacch
15	145.5	3.5	1240	YXN1_YEAST	P53935 saccharomyc
16	143.5	3.4	900	SYA_MYCGE	P47534 mycoplasma
17	142.5	3.4	875	Z1P1_YEAST	P31111 saccharomyc
18	142.5	3.4	1066	SPS5_SCHPO	O74451 schizosacch
19	142	3.4	850	D7_DICDI	P54682 dictyostell
20	142	3.4	1744	TANA_XENLA	Q01505 xenopus lae
21	141.5	3.4	490	MOT3_YEAST	P54785 saccharomyc
22	141.5	3.4	1928	MYSL1_YEAST	P08964 saccharomyc
23	141.5	3.4	2314	PRP2_HUMAN	P23471 homo sapien
24	141.5	3.4	2748	NIM1_YEAST	Q00402 saccharomyc
25	141	3.4	914	IF42_YEAST	P33936 saccharomyc
26	141	3.4	1332	SPF7_YEAST	P35177 saccharomyc
27	140.5	3.4	516	P54_ENTFC	P13692 enterococcu
28	139.5	3.3	1609	LMG1_HUMAN	P11047 homo sapien
29	139	3.3	1957	YDB6_SCHPO	Q10401 schizosacch
30	138.5	3.3	1703	SNF2_YEAST	P22082 saccharomyc
31	138	3.3	1639	MSF1_PLAEM	P04933 plasmodium
32	137.5	3.3	1790	USO1_YEAST	P23386 saccharomyc
33	137.5	3.3	2175	HMCU_DROME	P10180 drosophila

34	137.5	3.3	2459	1	MAPB_RAT	P15205 rattus norv
35	137	3.3	1176	1	PTNL_MOUSE	O62136 mus musculu
36	136.5	3.3	679	1	YIS3_YEAST	P40563 saccharomyc
37	136.5	3.3	1183	1	CNA_STRAU	O53654 staphylococ
38	136	3.3	1630	1	MSF1_PLAEM	P04932 plasmodium
39	136	3.3	1636	1	BUD3_YEAST	P25558 saccharomyc
40	135.5	3.3	1053	1	SLPM_BACBR	P06546 bacillus br
41	135.5	3.3	1976	1	MYHA_BOVIN	Q27991 bos taurus
42	135.5	3.3	3358	1	PGCV_MOUSE	O62059 mus musculu
43	135	3.2	1532	1	IGA_NEIGO	P09790 neisseria g
44	134.5	3.2	960	1	YMX6_YEAST	Q04279 saccharomyc
45	134.5	3.2	1124	1	TCF8_HUMAN	P37275 homo sapien

ALIGNMENTS

```
RESULT 1
ID YF16_YEAST
AC P43597:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE Hypothetical 137.7 kDa protein in UGSI-FAB1 intergenic region.
GN YFR016C.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95400292; PubMed=7670463;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae.";
RL Nat. Genet. 10:261-268(1995).
CC -1 SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: D50617; BAA09255.1; -.
DR SGD: S0001912; YFR016C.
DR KJ Hypothetical protein.
DR KW
DR SQ
DR SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;
Query Match 4.1%; Score 171; DB 1; Length 1233;
Best Local Similarity 19.8%; Pred No. 0.1; Mismatches 402; Indels 222; Gaps 42;
Matches 189; Conservative 142;
```

QY	12	VKENRVSYIDKQATKTENTLPDEVSKEGINAEQIY-----KIQD-QGYVTS 61
DB	151	IKETSNVNAEG-----TENVP--IKESTGIEVGNSTPRKKKKKKKTTNRGRNS 202
QY	62	HDDHHYHYNGKVPYDAIT--SELLMKDPNLYLKDEDI-----VNEVGGYIYKD 110
DB	203	NPADTDTDSKOSTLSDISGIEVYLOEDGS---KNEDIVNTVODEPVAVEMDRIITRNE 259
QY	111	GRYYVYLKDAADANV-----RTKEINRKOEHSHRGSG--TPRNDGVALARSQ--- 160
DB	260	SSDKFPDIDVPKNDVDESSKSENNINNEKAHHTLPRENTILVNBGNAASFHOLEP 319
QY	161	GRYTTDDGYIFPNASDIETDGDAYIVPGDHYHYIIPKNELASASELAAEAFISGRGNLS 219

```

Db 320 HGLEAGD-----NGAOSTKDVSESLTKNGFNKESKHLKAGEKOQTE---SDRDGIS 372
OY 220 NSRTYRONSDMTSRTNWVPSVSNP-----GTTNTNTSNSTNSQASQNDI----- 267
Db 373 PSYLANQKETEETGKEDHFEQKDEKDEKCRKLNVNHNHNNMNNNAAGSDSITIPETE 432
OY 268 -----DSLKOLYKLPLOSRRV-----ESDGLVDPDPAQ---TISR 299
Db 433 RETYDEETMGPTKRISDNENKLNQHTNDISVEVEKEEEEEEEENSTFSKVENVTGE 492
OY 300 -----TARVNAVRPHGDHNFITYSOMSELEERIALIPIRYASN----- 358
Db 493 QEAVERNNEVSGTEESTSKGEELMGD-----EKQSEAGEK-SSITIEGSANSAKIS 544
OY 339 ---HWVDSRPEQSPQPTPEPSPQAPNPLKIDSNSLSQLYRKVYEGEYVEEKGISR 396
Db 545 KDLVLEDEAEAPTOEKKPTPEVYGEIDIPAPRD-----VELVEAVENKIIIPDELEVAK 599
OY 397 -----YVFAKDLPSETVKNIESKLSKQESVSHTLAKKENAVAR-----DOEFTDK 442
Db 600 EDQEGQVOKLDEPVKAMKDDKIMRGAESISEDMKKRQDETAALSNBKAKEVEDTARES 659
OY 443 AYVULTEAHKALPYXN-----GRNSDFQALDKLERLNDESTNKEKLVDDLAFIAP 495
Db 660 AEGVEVEKSKTPSPKPVYKRTSGRPEDLIQINERDPEVLAKEDVVRVDEDKKPEIATTIEN 719
OY 496 THERLGRPNQIETEYDEEYRIQALADKYTSDGYIFDEHD-----IISDEGDAYYTPHM 550
Db 720 SEEDBPCKSQOISTBOAETITQKMDGVGSTS---FKEEKKRPREITQEGKIKTGKDT 776
OY 551 GSHHWIGKDSLKKEVYAAQATYKKEGILPPS-----PDADVKNANPTGDSAAAIYNR 602
Db 777 NHHGATEAASSENSK-ASDVGTAEKYIETPSSSESVKKTDEDAEVE---NSEKTEFIK 830
OY 603 VKGEKRPILRYLVEVHEVTEVKNGLI-----PHKDYHNKIFAMPDHY 650
Db 831 VKAE--LENDADKAEVTAELNKENEDVEVTEEDAENVSSEKTEIKVK---AELGNL 885
OY 651 KAPN--GYTLE-----DLFATIRYVE---HDERPHSNDGNGMASEHVLGRKD-- 694
Db 386 DAKKAEVTAELNKENEDVAVATSKEDIETKSEPAEPIED---GTCEAAVSKDAE 942
OY 695 ---HSEDPNKKFADEBPVEETPAEPPEVPE-----TEKVE-AOLKAEVYLLAKYDSSL 746
Db 943 AATKEENMNSKIALKLVGTGDOEIDIDINISDEFORTVELPELKEKODIKDKGGEKEL 1002
OY 747 KAAATEPLAGL-----RNNLTQIMONNSIMAEKLLILKGSNSSSVSKEN 796
Db 1003 EYEETEKETSLPDLVYEENIT---EENKELKQEEB-VSOLDENETESISKAPN 1053

```

RESULT 2  
APX\_XENLA  
ID APX\_XENLA STANDARD: PRT: 1420 AA.

AC 001613;  
DT 01-JUN-1994 (Rel. 29, Created)  
DT 01-JUN-1994 (Rel. 29, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE Apical protein (APX).  
GN APX.  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Ovary;  
MEDLINE=93107151; PubMed=1334959;  
RA Staub O., Verrey F., Kleyman T.R., Benos D.J., Rossier B.C.,  
RA Kraehenbuhl J.-P.,  
RT "Primary structure of an apical protein from Xenopus laevis that  
participates in amiloride-sensitive sodium channel activity.";

```

RL J. Cell Biol. 119:1497-1506(1992).
CC -I- FUNCTION: IS PART OF A MULTIMERIC COMPLEX WHICH IS INVOLVED IN
CC AMILORIDE-SENSITIVE SODIUM CHANNEL ACTIVITY.
CC -I- SUBCELLULAR LOCATION: Membrane-associated.
CC -I- TISSUE SPECIFICITY: KIDNEY, PROXIMAL INTESTINE, OOCYTES, AND
CC TO A LESSER EXTENT IN THE DISTAL INTESTINE, STOMACH AND EYE.
CC -I- SIMILARITY: SOME, TO HUMAN APXL.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: Z14997; CAA78718.1; -.
DR PIR: S25517; S25517.
KW Membrane; Sodium transport; Transport.
FT DOMAIN 111 114 POLY-SER.
FT DOMAIN 559 564 POLY-SER.
FT DOMAIN 1048 1051 POLY-THR.
SQ SEQUENCE 1420 AA; 159467 MM; 0DD8B5C11413FBFC CRC64;

Query Match 3.9%; Score 161; DB 1; Length 1420;
Best Local Similarity 19.8%; Pred. No. 0.42;
Matches 180; Conservative 111; Mismatches 318; Indels 302; Gaps 44;

OY 42 EGINAQIYIKITDQGYVSHGDY-----HYNCKVYV-----DAISFE 82
Db 26 ERISPVRSMTTLVDSAYSSFSGSYVEYONSFQHDGCHYNDQLSYMDSEYVALIYNS 85
OY 83 LMKDQPNYKIKDEDIVNEKGVYIKVDK-----YYUULDAHADNVRTK-- 129
Db 86 LLDKQGVY---NDIYSE-HGSSKVALSGRSSSLCSDDMTSVHRKSPAKLDNYVNLDS 140
OY 130 -----EELN-RQKQESHQHREGGPRND--GAVALARSGR-----YTTDDGYIFNA 173
Db 141 EKNYGDPINMKHKQRPNHKAYGLQNRSPGTGINSIQEKENDQLYNPSNMEIKDNFYGRS 200
OY 174 SDLIETDGAIVPHGDHN-----YIPKNELASLSLAALAEFLSGNGLSNR-----TYR 225
Db 201 LDVLQDGD--IMTQDSYTONALYFPQNPQOYR---NTQYPGANRSKQOFKYNDVQ 253
OY 226 RONSNDTSRTNWPVSVSNGTNTNTNSNNSNTNSQASQNSDIDSLKOLYKLPLOSRAVE 285
Db 254 KSNENETERDG--PYITKQGF-----VQGYASDVRTSKNI-----RSLK 294
OY 286 SDGLVDPQAQITSRTARG---VAVPRGDYHF-----IPYSOMSELEERIALI-- 331
Db 295 KSA-----SGKIYVHDQSGCWMKPKGOTPFNFSEGTITDMOYDNREQDIRKSLSTRA 350
OY 332 --PLRYRSNHWVDSRPEQSPQPTPEPSPQAPNPLKIDSNSLSVQLYRKVYEGGYV 389
Db 351 SOSLYYESMEDV-----SGPPLKAMMSKNVDLTLS----- 381
OY 390 EEKGISRYFAKDLPSETVKNIESKLSKQESVSHTLAKKENAVAPRDOEYDKA-----Y 444
Db 426 HLAGHSAFIAVHNTPNAQOEKLEKLETKLERMNNIS-----VLQISEPRDNH 476
OY 500 RUGKPNQIETEYDEEYRIQALADKYTSDGYIFDEHDIISDQDAVYTPHMGSHWIGKD 559
Db 477 KLPKNNKS-----LTQLADLHDSVEG-----GNSGNLNSABE 508
OY 560 SLS-----DKEVYAAQATYKREKGI---LPPSPDADVKNAP----- 591
Db 509 SLKNDYIEKLVAAQKAVLRETSFKRKDLQMSLPCRFKLPKRPRTIDHFRSYSSSSANDE 568

```



```

OY 592 -- --TCGSAALY-----NRVGEKRIPLVRLPYMEHVEHYVKNGLIIPHK-D 635
Db 569 SAYLQTKNSADSYKKDDTEKYAVTRIGCRKI-----TRQOKKLCYSEPKID 617
OY 636 H--YHNINFAWEDDHTYKAPNGCTLEDLFATIKYVVEHPDERPHSNDGMSNASEHYLGRK 693
Db 618 HLGIGKSNFAWKEEPTFANRREMDSDDISANRIKYLE-SKEVTNNS---SNLSKTELRQI 673
OY 694 DH-----SDDPKNFKADEPEVEEPA-----EPPEVPEYETEKAQLK----- 732
Db 674 OHNALVOYMERKTNORPNSN--PQVOMERTSLGLPNYNEMWISYSETSSDASQKYLR 730
OY 733 ---DAEVLAKVT-----DSSIKANATETLAGLRNNLTQIINDNNSIMAEKLLAL 781
Db 731 RSAGSSSYDATALVTMNDRFGRKTSPLGRSNAEKTAGVQK---TFSDQRTLDQSQEH----- 783
OY 782 LKGSNPSSVSK 792
Db 784 LEGSSPSLSQK 794

RESULT 3
UBP8_HUMAN
ID_UBP8_HUMAN STANDARD; PRT: 1118 AA.
AC P40818;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 8) (Ubiquitin-specific processing protease 8)
DE USP8 OR KIAA0055.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
CC -I- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/isb.ch).
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D29956; BAA06225.1; -.
DR MEROPS: C19.011; -.
DR GeneW: HGNC:12631; USP8.
DR MIM: 603158; -.
DR InterPro: IPR0011763; Rhodanese-like.
DR InterPro: IPR001394; UCH-2.
DR Pfam: PF00442; UCH-1; 1.
DR Pfam: PF00443; UCH-2; 1.
DR Pfam: PF00581; Rhodanese; 1.
DR SMART: SM00450; RHOD: 1.
DR PROSITE: PS00972; UCH_2_1; 1.
DR PROSITE: PS00973; UCH_2_2; 1.
DR PROSITE: PS50235; UCH_2_3; 1.
DR Ubl conjugation pathway: Hydrolase; Thiol protease; Multigene family.
FW ACT SITE 786
TX BY SIMILARITY.

```

FT	ACT_SITE	1059	1059	BY SIMILARITY.
FT	ACT_SITE	1067	1067	BY SIMILARITY.
SQ	SEQUENCE	1118	AA; 127523	MM; 8B884B7A842FE99A
				CRC64;
Query Match		3.8%;	Score 158;	DB 1; Length 1118;
Best Local Similarity		19.0%;	Pred. NO. 0.44;	
Matches 173;	Conservative	143;	Mismatches	336; Indels 256; Gaps 455;

QY	6	LYQANTVEENNRYSTIIOGKATQKTENLTPEDEVSAREGINAEQVIKITYITDOGYTSHGDH	65
Db	11	LYLSLSSLDLUNK-----KTE-VKPEISTKSYSHALIKFYKAECCRLDR--DE	56
QY	66	YHYNGKPYOALIISEELMKDPNKLDEDIIVNKGGYIKYKDGKYYVYLKQAA-----	122
Db	57	ERAYLYKYYTYVIN--LTKRPDEK-OQODFHSILBPGNIKAVEABERLSCLKRY	111
QY	122	HADNVRTK-EELINROKO-EHSOHREGCTPRNDGAVAL-----	156
Db	114	EEAEVRKLLKEEDRQEEAORLOQRKQETGRDGGLAKGSLNVLDSDKTKOKSNGEKNE	177
QY	157	---ARSQRYTDDGYIF-----NASDII-----EDTGDAIVPHGDHYHYIPKNELAS	203
Db	174	KCEETKEKALITAKELYTMATDKNISLIIIMDARRMODYODSCIL-----HSLSPVEEALISPG	222
QY	204	ELAA-AEAFLSGRGLMSRTYRRQNSDNTSRTIMWVSVSN--PGTNTNTSNNSNTNSQ	266
Db	230	VTASITAEHLR---DDSKOTMKKRWENYVLLWLFSSAKLOIGTTLRLSLK-----	278
QY	261	ASQSNIDISLLKOLYKLPILSORHESD-----LVFPDAQITSTRAGVAVPHGDHYFI	315
Db	279	-----DALRKMEKTYLARNBPLVLEGEYENMLLCYQYTTNKK--VYPP-----	322
QY	316	PYSQMSLEEFRIARTIPLRTYSNMHWPDSPREQSPQPTPEPSPGQOPAPMLKIDSNSL	375
Db	321	PRRQ---NEEVSISLDTYPS--LEBSIPSKPAAQ-----TPPASIEVDENIEL	364
QY	376	VSQLRKVG-----EGYVEEKGISRYVAFANDLPSEYTKNL-----ES	413
Db	385	ISGQNERGPLNISTEYEPVPAASSDVSPTI-----QRPASIKNPQIDRTKKPAVCLPEE	420
QY	414	KLSQESVSHTLTKKKEVNAVRDQDEFYDKAYN---LLETAHK-----ALFXNKGNSD	463
Db	421	HRKSESINHHQSQSQSQSKVLPDSTKRVVSPPLMLLDEKARIHAETALMLKKNQE-	479
QY	464	FOALDKLIERLNDSTYNEKELVDDLAFIADITPERLGRFNSOI---EYTEDVIRIQAOL	520
Db	480	---KELRBOOE--QKEKLRKE-----BOEQAKKQKQAEHEBELTEKQOKAKEE	524
QY	521	ADKTYTSGYIFDEHDIISDGDAAVYTPPHMHSWIGKDISDKENYAAQAQYTERKILP	584
Db	525	MEKKESEQAKKEDKE--TSARKGEIT-----GVKRQSSSEHETSDAKSVYDGRKRC	575
QY	581	PSPDADYVA-----NPTQGSAAIYIRKVGKRIPLVLRPLYVWEHTVEVKNGNL----	629
Db	576	PTPELOKKSSTDYDHTSVTSGDSSGSKPFKIKGQESGLIRKGTRETDOTERKKAOREP	635
QY	630	-----IIPKHDIYHNIKAFMFDDTYKAPNGYT--LEDLFATIKYY-----VEH	671
Db	636	LTRARSEEMGRIVP-----GLPSGMAKFLDITGTFRYHSHPTNTVH	677
QY	672	-----PDERPHSNDGKMSSEHYVLGKKHSDPPNNKFKADEEP-----VEETPAPEE	718
Db	678	MYPPMASSAPSPSTPTTHKAKPOIAPERD--REPSKIKRYSPPDITQALQIEEKKRPT	735
QY	719	V-POVETEKVPAOLKEAEVLLAKITDLSLK-----ANATETLAGLRN-----NLTIQ	764
Db	736	VTPIYNRENRKCTCPKAEI--SRLSAQIRNLNVNPFQSSGALLGLANLGNCTYMNISIQ	793
QY	765	IMDNNSIMAE 774	
Db	794	CICNAPHIAD 803	

```

RESULT 4
ID DNM_MYCPN STANDARD; PRT; 910 AA.
AC P75354;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaU-like protein MG200 homolog.
GN MPN19 OR MP035.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RA MEDLINE=97105885; PubMed=8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.,
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae."
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -I- SIMILARITY: CONTAINS 1 J DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AE000004; AAB95683.1; -.
DR HSSP; P25685; 1HDJ.
DR InterPro; IPR001623; DnaU_N.
DR Pfam; PF00226; DnaU_1.
DR SMART; SM00271; DnaU_1.
DR PROSITE; PS00636; DnaU_1; 1.
DR PROSITE; PS50076; DnaU_2; 1.
KW Hypothetical protein; Chapterone; Complete proteome.
FT DOMAIN 4 73
FT SEQUENCE 910 AA; 100190 MW; 125D0E37D2D221A7 CRC64;
SQ
Query Match 3.8%; Score 157; DB 1; Length 910;
Best Local Similarity 19.0%; Pred. No. 0.38;
Matches 161; Conservative 127; Mismatches 305; Indels 254; Gaps 44;
OY 66 YHYNGKGVYDIIISEL-----LMKP---NYKLKDEIYNEVGVIKVD--GKTY 114
DB 34 YHDKRKAPDAAOIFAEIENADVLSNPKRKANYDKYGHGVDN-EGGFAFOADVDFSE 92
OY 115 VYLKDAHADNVRTKEINROKQESHQREGTPRNDGAVALARSGRTTDDGYIFNNS 174
DB 93 EEIEKSGARDNL---SESTKKKEKTKTKKGW-----FMGSKQDEESTSTTEY--A 140
OY 175 DIIETGDGYIYPHGCHYIYIPKNEISASELAAEFLSGRLNSKRYRRONSNTSR 234
DB 141 DV--DAGLEDYPPQSDYPPDIPDVARIIEVQS--AYADIPDVAGMDEQNAEYANSA 197
OY 235 TMMVPSVSPGTTNTNTNSNTNSQASQSDIDSLL----- 271
DB 198 SELIIPV-----DAGLAEFTNTSSAQAQSDWEANTGMPYCYFPDAGEWMKKGFEDA 251
OY 272 -KQYKRLPLSORHVESDGLVFPDPAQTSRTARGVAVPHGDHYHFIYPSOMSELEERARI 330
DB 252 GQVMVLEETPESSVSNDFTTSDAVTAAT-----VEETIDQ----- 288
OY 331 IFLKRNINWVDSRE-----QPSQPTP-----EPSPQP-----QAPNKKI 369
DB 289 -----DSWTNNSAPEPVAVETPELQPEPEPIITLTSSEPVAPASVIEPIPEIBE 341
OY 370 DSNSSLVQLVARKVGVYEEKSGISRYVFAKDLSEYTKNLESKSKSESHITLAK 429
DB 342 TTSAVEMDAVSK----ADVSDADATNPTEDDTISEPQ--ETDAALAEIINT--TADL 394

```

```

OY 430 ENVAPROEYDKAVNLTLE--AHKALFXNKGNSDFOALDKLERLNDSTNKEKLYVD 487
DB 395 E--PAEV-----SATNDLQDVYEKVFSEPESTVDTATDPVDAQETTSNGFFFE-N 446
OY 488 LIAFL-----APITPERLGRKPNISOI-----EYDEYRIQOLA----- 521
DB 447 FSSFVLSIDQNPNPQPTTHHEEDAAPETVDETSGETSTPEVTIASTVELETAIEINNP 506
OY 522 -----DKYTSIDGYIFDEHDIISDEGDAVVPFHMGSHWIGKDSLS 562
DB 507 AIFVEEYLOPTKTYVDKLDPELPAKPTVSDSENSVAPPEPVACGEQTF--W--KPAIS 563
OY 563 DKEKV-----AAQAYTKKKGILPPSPDA-----DYKANPTGDSAAA-----IY 600
DB 564 EEHEIPLTAVEPASETQTLIAEDVTSVPVPTAFAPISINAVPAPVATEEAADFELK 623
OY 601 NRKGEKRIPLV-RLEPMVEHTEYKNGNLIIPHKHDTYNIKFPAMDHTFYAPNGCYTLE 659
DB 624 EAKIEAQLPLVPTVEQIDGT-----DPSIL-----TQWDEYLEKTRK--LPH 665
OY 660 DLFATIKYVEHDERPHSHNDGKGNASEHVLGKDHSE--DPNKNKKADEEVEETPAEP 717
DB 666 KLFLL-----EQLP-----FIVKTDQFEIYDPA-----LDEHVNLLITE- 700
OY 718 EYPOVETEKEVQAQLKEAEVLLAKVTDSLSKANATETLAQLNNLTLOIMDNNSINAEAK 777
DB 701 HVPQICF--LNGQLKIR-YTRKLVDPQVYTTT-----SITLEVLGSHKSQTEA-- 748
OY 778 LIALKLG 784
DB 749 -IAIFKG 754

```

```

RESULT 5
ID Y67_YEAST STANDARD; PRT; 1658 AA.
AC 003661; Q04988;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 187.1 kDa protein in GUA1-Eng8 intergenic region.
GN YMR219W OR YMR261.13 OR YMR959.01.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Saccharomycetes; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE OF 1-711 FROM N.A.
RC STRAIN=S288C / AB972;
RA Dedman K., Brown D., Bowman S., Barrell B.G., Rajandream M.A.,
RA Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 608-1648 FROM N.A.
RC STRAIN=S288C / AB972;
RA Skelton J., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: Z49809; CAAB9934.1; -.
DR EMBL; Z49939; CAAG0190.1; -.
DR SGD; S0004832; YMR219W.
KW Hypothetical protein.
SQ SEQUENCE 1658 AA; 187137 MW; 3893F968305A575D CRC64;

```



```

Db 29 HSEBPLEPESTINSS-----FNDSVNTESDIASKSD-VPVS-----S 68
OY 248 NNTS-----NNSNTSQAOSANDI-----DSLKOLYKLP 279
Db 69 SNINSPANETOLEIPDTOLHKLKLANDSDQHDITADSDNLPNSIYEHDSVITOT-KPAM 127
OY 280 SORHVESDGLVDPQAITSR-----TARGVAAPHGDHFIPIYSOMSE----- 322
Db 128 SOYEET-----AAHLSRNPISLDVAGELHNNHNTOKIAVSAVEEDSFNEEGENDH 181
OY 323 --LEERIAITILKRSNNHWPDSRPEQSPPTPEPSGPAPANLKIDSNSLSVQLV 380
Db 182 STIISLNDATSOY--NHFLPSD-----GNLSPELSSGDTPTNHVPLGTNDN----- 228
OY 381 RYVGEYEEKGRIRYVAKD-LPSEYVKNESKLSKQESYSHLTAKENAVPPODF 439
Db 229 -EINDEYONDEIS--LNANVLDELDELKEDERKLETHYS--TEEEKODIA--DOET 281
OY 440 YKAVNLTLEAHKALEFXNKRNSD--FOALDKLLERLNDELSTNKEKVLADLLAFLAPITH 497
Db 282 AENLTSTSTPESENKIRNSGDDTSMLFQ-----DDESQKVPWEDVKDF-----H 328
OY 498 PERLKPNQIETDEVR-----IAQLADKYTSDGIIPDE---HDIISDEGAYVTP 548
Db 329 NENTNTOESAPNTDROKGYEGNEALKKSECTADERSYSEETSEDFHGHDKQVVE- 387
OY 549 HNGHSHWICKDSLDPKEKVAQAATYKEKILPSPADAV--KANPTGDSAAAIYNNKGE 606
Db 388 --GONDFCKNINENESQKLMGEGNHK---LPLSADIIIEGKODQDAEDLFTQSSD 441
OY 607 KRIPLVRLPYVNEHTVEYKGNLIIIPKDHYNI-----KFWN---FDDHYKADN 654
Db 442 -----LGEVLPWMESTDNADVTSKSCKEHLDFASGNDKLPWESVDSGEVSSGKTEN 494
OY 655 G-YTLEDLFAFIKY-VYEHDERPHSNDGMSNAHEVGLKHSDEPNKFNKADEPVEE 712
Db 495 SNOTSTERTAECKFSFLENDDDLDDDSFLASSE---EEDTVPNTDNTJLTSKPVVE 550
OY 713 TPAREPPOVETKEVAQAEVLLAKVT 742
Db 551 KKASRYKPIIEE---EAGMRQOVHHTNTT 577

RESULT 7
MAPX_DROME STANDARD: PRT; 1185 AA.
AC P23226: 09V9S1;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 205 kDa microtubule-associated protein.
GN MAP205 OR CG1483.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91115949; PubMed=1703540;
RA Irminger-Finger I., Laymon R.A., Goldstein L.S.B.;
RT "Analysis of the primary sequence and microtubule-binding region of
RL the Drosophila 205k MAP."
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM B3).
RX STRAIN=Berkley;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortens Y.-H.C., Blazew J.R.G., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers J.-H.C., Blazew J.R.G., Champe M., Pfeiffer B.D.,

```

```

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Crawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaitlin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimms I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svrtkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE REGULATION OF
CC MICROTUBULE ASSEMBLY AND INTERACTION.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH CYTOPLASMIC MICROTUBULES AND
CC WITH THE MITOTIC SPINDLE.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; C2, B3 AND J5 (SHOWN HERE); ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- MISCELLANEOUS: PHOSPHORYLATION OF VARIOUS SERINE RESIDUES MAY PLAY
CC A REGULATORY ROLE. THE BASIC DOMAIN CONTAINS NUMEROUS SEQUENCES
CC THAT MATCH KNOWN CONSENSUS SEQUENCES OF SEVERAL DIFFERENT PROTEIN
CC KINASES.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X54061; CAA37996.1; -.
DR EMBL: AE003780; AAF57214.1; -.
DR PIR: A36685; A36685.
DR Flybase: FBgn002645; Map205.
KW Microtubules; Alternative splicing; Phosphorylation.
FT DOMAIN 1 784
FT DOMAIN 785 1124
FT BINDING 1125 1185 ASP/GLU-RICH (BASIC).
FT BINDING 745 977 ASP/GLU-RICH (ACIDIC).
FT VARSPLIC 557 578 TO MICROTUBULES (POTENTIAL).
FT VARSPLIC 650 703 MISSING (IN ISOFORM B3 AND ISOFORM C2).
FT VARSPLIC 704 704 MISSING (IN ISOFORM C2).
FT SEQUENCE 1185 AA; 126669 MM; 4784222ECED03F70 CRC64;
SO Query Match 3.8%; Score 156.5; DB 1; Length 1185;
Best Local Similarity 20.1%; Pred. No. 0.58;
Matches 163; Conservative 111; Mismatches 304; Indels 231; Gaps 36;
OY 122 HADNVRTKEI-NRQKQ-----EHSQHREGGTPRNDGAVVALARSGRITDGGYIF 171
Db 4 HEDNOLDNYLONRLAESIQICGAGEHNPHLADATGGNGCAPGAPAPSK----- 52

```



```
FT REPEAT 232 261 ANK 6.
FT REPEAT 265 294 ANK 7.
FT REPEAT 298 327 ANK 8.
FT REPEAT 331 360 ANK 9.
FT REPEAT 364 393 ANK 10.
FT REPEAT 397 426 ANK 11.
FT REPEAT 430 459 ANK 12.
FT REPEAT 463 492 ANK 13.
FT REPEAT 496 525 ANK 14.
FT REPEAT 529 558 ANK 15.
FT REPEAT 562 591 ANK 16.
FT REPEAT 595 624 ANK 17.
FT REPEAT 628 657 ANK 18.
FT REPEAT 661 690 ANK 19.
FT REPEAT 694 723 ANK 20.
FT REPEAT 727 756 ANK 21.
FT REPEAT 760 789 ANK 22.
FT REPEAT 793 822 ANK 23.
FT DOMAIN 1773 1950 REPEAT-RICH REGION.
FT REPEAT 1773 1784 REPEAT A.
FT REPEAT 1785 1796 REPEAT A.
FT REPEAT 1797 1808 REPEAT A.
FT REPEAT 1809 1820 REPEAT A.
FT REPEAT 1821 1832 REPEAT A.
FT REPEAT 1833 1844 REPEAT A.
FT REPEAT 1845 1856 REPEAT A.
FT REPEAT 1857 1867 REPEAT A.
FT REPEAT 1868 1879 REPEAT A.
FT REPEAT 1880 1891 REPEAT A.
FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).
FT REPEAT 1903 1914 REPEAT A.
FT REPEAT 1915 1926 REPEAT A.
FT REPEAT 1927 1938 REPEAT A.
FT REPEAT 1939 1950 REPEAT A.
FT DOMAIN 3536 3620 DEATH.
FT VANSPLIC 1039 1039 Q -> Q1GKHLHPAPPPLGESLVSRIILQLEPPGPK
      (IN ISOFORM 2).
FT VANSPLIC 1444 3528 MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT CONFLICT 475 476 GO -> PE (IN REF. 4).
FT CONFLICT 971 971 I -> S (IN REF. 1).
FT CONFLICT 3581 3582 QY -> HA (IN REF. 1).
FT CONFLICT 3586 3586 I -> Y (IN REF. 1).
SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C428E29D2 CRC64;

Query Match      3.7%; Score 154.5; DB 1; Length 3924;
Best Local Similarity 18.0%; Pred. No. 3.8;
Matches 164; Conservative 128; Mismatches 232; Indels 387; Gaps 45;
```

```
Db 3275 TMTSTPAD-----PSAEYESS-----VSEDFLSVDEENKADKPKSLPVKV 3319
QY 408 -VKNLSEKSKQESVSHITLAKKENVAPRDQFYDKAYLLLEAKHKLFXKNGRN-SDQ 465
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3320 PLQRYEQQLSDLD-----TSVQKTVAPQGDMASTAPD-----NKSSESQ 3362
QY 466 ALD-----KLERLNDESTNKEKLVLDLAFAP1-----495
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3363 SLDKTKCPVKTRSTYETETESRERAELESEBQATPKILTRSL-----PKRSRTS 3418
QY 496 -----THPERLQPN-----SOLEYTE-----DEVRIAQLADKYTTSQGIYFDEHDIISDE 541
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3419 SCRGSTPTKESKEHEFFDLYRNSIEFFEFEISDEA--SKLYDRLTQSE-----REQETVSDD 3472
QY 542 GDAYTTPHMGSHHWIGKDSLSDKEKVAQAAYTKKGIILPPSPDAVKANPITGDSAAIYN 601
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3473 -----ESSAL-- 3478
QY 602 RVKGEKRIPLVRLPYMVEHTVEVKNQNLIIIPKDHYNIKFAMPDDHYKAPNGYTLDEL 661
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3479 EVSVIENLP-----PVETHSVP-----EDI 3499
QY 662 FAT-----IKYVEH-PDERPHSNDGMASEHYLGKDHSEDPNNKFADEEPEEET 713
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3500 FDTRLPMDESIETLIERIPDENG-----DHAEDP-----QDEORIEER 3539
QY 714 PA-----EPEVPOVETEKVQAQLKEAEVLL-----AVTSS 745
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3540 IAYIDHIGFSTELAREIDFTEEOIHQIRIENPNSLODOSYLILKIWERDKIATDIN 3599
QY 746 L-----KANA-----TETL-----AGLRNNLTLOIMNNSIMAEKLLALL 782
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3600 IVECLTKIRMDIVHMETNTEPLQERISHSYAELEQITLIDHSGEFVLOE-ELCTNOH 3658
QY 783 KGSNSSVSKE 793
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3659 KQKEQAVSKE 3669

RESULT 9
Y373_BOVIN
ID Y373_BOVIN STANDARD: PRT; 1453 AA.
AC Q9TU23;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypoetical protein KIAA0373.
GN KIAA0373.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Trachea;
RA Jovov B., Ripoll P.J., Benos D.J.;
RL Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF176816; AAF00990.1; -
CC DR Hypoetical protein: Coiled coil.
KW DOMAIN 37 1426
SQ SEQUENCE 1453 AA; 169934 MW; 03CBA02A64CF4139 CRC64;
```

Query Match	3.7%	Score 154	DB 1	Length 1453
Best Local Similarity	18.5%	Pred. No. 1		
Matches 171	Conservative 124	Mismatches 329	Indels 298	Gaps 40
OY	7	YQAFVFKENNRVSYIDGQA-----	TKTENTL---	PD-----EYSKREGINAEQI 49
Db	486	WQREVELEERQUDVDRROOSELRPAQKFEAYGSGMPDPSLIPMOLEIALKK--	IKETI 543	
OY	50	VIKITDQGVYTHSGDHNYHNGKVPYDAIISSELLMKDPNYKLAKDEDI-----	YNEVG 103	
Db	544	RILETQATCRS-----	LEKLAKKEBSALRLAEENISLRKYINELR- 585	
OY	104	GYIVVDGKYVYLYLKDAAHADVNTKKEELINROKQESHOSHREGGTPRNDGAVALAASQGRY 163		
Db	586	-----	LRLPATAEQEKLLAEFSRKEVEPKSHHTLKLAHQ-----TIAMQARL 628	
OY	164	TTDDGIYFNASDIIEDTGD--	AYIVPHDHHYI--PKNELSASELAAAEAPLISGRNLS 219	
Db	629	NQKEEVLKKYOHLLKAREQREIYKKEHEELHTLRKLELOA-----	671	
OY	220	NSRTYRQNSDHT-----SRTNW--	VPYSNPGTNTNTNSNNSN--TNSQASQSDIDSL- 270	
Db	672	-----	DNSLSKFETAMDLIKQSEPTPYPTKKHFLRLAEMQGTAAEQDSDSLSVI 721	
OY	271	LKOL-----	YKLPLSORH-----	VSDGIYVPPAQITSR 299
Db	722	KIKGYASODLERQKETEELKEIKFEENMKMLRLQGNHNADEVYKIKAEVYEDLRCLLYVQSGESQ 781		
OY	300	TARGVAVPHGDHYHTIPYSQNSELEERTARITPLRYRS--	NHWVPDSRPQSPQPTP 355	
Db	782	SLKSELQTOKEANSRAPTTMRLNVERLKSQALALEKQOKALSRALLERAMTAAEER 841		
OY	356	BSPSPQAPAPAKIDS-----	NSLVSQVLRKVGSGYVEEKGISRYAFKAPLSEETYN 410	
Db	842	IISMTSQEAMLNVOQIYDRHTKEKSOI-----	EDL--NENILK 879	
OY	411	LESKLSKQESVSHTLTAKKENVAPRDOEYFD--	KAYNLLTEAHKALFYXKGRNSFOALD 468	
Db	880	LKEALTKTSKNRENTLT---	DNLNDLTNELQNKQAYG-----KYLREKDAVDQENNELK 930	
OY	469	KILLETLANDE-----	STNKEKLYVDLLAFAPITPERLGRKNSQIETTEDEVRLAQIAD 522	
Db	931	QKIKRLTSGLOGKPLIDNKQSLIEEL-----	QKIKIKLESOLERKRVDEAKKPMKE 981	
OY	523	KYTTSDGYIFDE-----	HDIIISDEGDVAVTTPMHGSHMIGG--	SLSDKEVYA 568
Db	982	KSAREEVIRWEGGKKWQRTIGISIRNKKLKEKESEVYI--	LTQOLTTLKDLFEAKDEKYL 1038	
OY	569	AQAVYKEKGILPPSPDADYKANPTGDSAAAIYNRKGEKRIPLVRLPYVEHTVEYKNGN 628		
Db	1039	LQRLKLTGTGL-----	TYDQWMAA--RVLESEKELE-----	BLKRN 1072
OY	629	LIIPIKDHYNIKFWMFDQHTYKAPNGTYLLEDLFATIKYYE--	HPDERPHSNDGWMAS 686	
Db	1073	LDLENDISYMRSHQA-----	LRDSYIEDLHQNXYLOERKTHALEKQSLSDAYSRPS 1124	
OY	687	EHVLGKKKH---	SSDPKKNFKADEPEY---ETPAPPEYVQVET-----	724
Db	1125	TSGIDSDHYOREQELQREBNELKSESENIELKFQLEQANKDIPRLKNQYVRLEKMECFIHK 1184		
OY	725	EKVEAQLKEAEV-----	LLAKYTTDSILKAN-----	ATEITLAG 756
Db	1185	EKAEEVRLTGRVRSGRSGKTIPELEKTIIGLKKYVKEVQRNEDQLKASGILTSEKMAN 1244		
OY	757	LRRNLTLOIMNNSIMAEKYL 778		
Db	1245	IE-----	MENEKLIKAELEKL 1259	
RESULT 10				
K373--HUMAN				
AC Y373--HUMAN	STANDARD;	PRT;	1539	AA.
AD O15078;				

```

DT      16-Oct-2001 (Rel. 40, Created)
DT      16-Oct-2001 (Rel. 40, Last sequence update)
DT      16-Oct-2001 (Rel. 40, Last annotation update)
DE      Hypothetical protein KIAA0373.
GN      KIAA0373.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Brain.
RX      MEDLINE=97349984; PubMed=9205641;
RA      Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA      Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT      "Prediction of the coding sequences of unidentified human genes. VII.
RT      The complete sequences of 100 new cDNA clones from brain which can
RT      code for large proteins in vitro.";
RL      DNA Res. 4:141-150(1997).
CC      -----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: AB002371; BAA20828.1; -
KW      Hypothetical protein; Coiled coil.
FT      DOMAIN             18 1514
SQ      SEQUENCE   1539 AA; 180065 MW; D90J314E981BF001 CRC64;

Query Match          3.7%; Score 153.5; DB 1; Length 1539;
Best Local Similarity 19.2%; Pred. No. 1.2;
Matches 200; Conservative 145; Mismatches 334; Indels 361; Gaps 53;

OY      3 ELGIQARTVKENNNVSTIDGKATQKTEN-----LTPEVSKREGINAEQIVIKITD 55
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      306 EQALYYAR-LIEGRNNAKH.-RQTQSLSRFSGALPDAQQEFSEKTMIOLODNKLKTM- 361
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      56 QGYVSHGDHYHVGKGPYDAIIIS--BELT--MKDP-----NYUKLKDEDI-VNEYKG 103
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      362 QEMKSQOEHRMEKKETLEMELKLGLEIISTLYKDYGAKGVITYMNHKITELRIQDELK 421
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      104 GYVIKVDKRYUYLLDKDAHADVRT---KEEINROKQEHNSQHGEGTPRRDGAVALAR 159
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      422 NRELTKDXKEIKILNNII-SEYEPTISSLEEIYQGNFNEERQAWMPRE---VDLERQ 477
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      160 OGRTYTDGCIYFNASDIIED-TGDAYIYPHGDDHYIYIRNEL-----            200
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      478 LDFPRQONEIINAQKFEEATGS--IP--DPSLPPL-NDEIALRKIKENIRIILETR 531
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      201 -----SASELAAEAFLSGRONLSNSTRPYRRONSNDTSRTMWVPVS-----NP 244
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      532 ATCKSLERKIKEESALLADEONIIISRKVINE--LRLLPATAREKLIAPELGKKEMP 589
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      245 GTTNT-NTSNNSNTNSQAQSQSDNDISLLKOLYLPLESQRHVESDGLVEDPAQITSRTARG 303
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      590 KSHHTLKLAHOTIAMQA-RLNQKEEVLYKKYORLLEKAREEQ-----RE 632
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      304 VAVPRGGDHVHFPRYSOMELDERLARIRPLRFRSHHWPPDSRPEDSPQPT----- 354
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      633 IVKKEEDLHIHHLELDQADSSL-----FKQTAW--DLAKOSPVTVPINKAFIRLAE 685
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      355 PEPSGPQAPNLKIDSNSLVSQLVR-----KVGEGVFE 390
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      686 MEQTVAEQD-----DSLSSILVKKKYVSODLERGETTELKVFKEFNKIQLQDENHDE 739
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
OY      391 EKGISRIYFAADL-----PSELVKKLESKL----- 415
           | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB      740 VKYVAAYE--EDULKYLDQSQEKSLSELDQAOKEANSRAPPTIMRLDYRLKLSQALK 797

```



```

Yy      416 -SKQSVSHTL-----TAKKE---NVAPRDQGFVFK-----          442
       ::::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      798 EKQQAISRALLELRAEKMTAAAEERIIATSSQEKHLNV---QQIVRHRRFELKTQVED 853
       |||::| | | | | | | | | | | | | | | | | | | | | | | | |
Yy      443 -AYNLTFAHKALFEFNKR-----NSDFQ---ALDKLL-----ERLND---    477
       |||::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      854 LNEENLL-KLKEKLTKSKRRENSLTQNLDINNELQKKQKAVKILIREKEEIQDENELKR 912
Yy      478 -----SYNKEKLVDDLAFIPIHBERLGKPNQSIEYTEDEYRIAQLADK 523
       :||::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      913 QIKRLTSGLOGKPLTDNKNQSLTEFL-----QRVKKLENLEGKVEVDLPKMEK 963
Yy      524 YTTSGCIYFDE-----HHIIDEGDAYYTPPMGSHHWIGKD--SLSDKEKAA 569
       ::::| | | | | | | | | | | | | | | | | | | | | | | | |
Db      964 NAKELLIMWEEKKKWAQAIIEGIIRNLKEREKGVPFLTRQNLTL--KDLFAKADKELKL 1020
Yy      570 QAATREKGLPSPDADYKANPTGSAAIYRVYGEKRIPLVRLPYMVEHTVEYKNCNL 629
Db      1021 QRKLKTTCG-----TVQVUGIT-RALESSELE-----ELKKRNL 1054
Yy      630 IIPKHDIHNLIKFAFDHDHYKAPNGYTLEDLFATIKYYVE--HPDERPHSNDGNGNSE 687
Db      1055 DLENDILVMRAHQ-----LPDSVAVEDHLGNRYLOEKTHALEKQFSKDTYSKPSI 1106
Yy      688 HVLGKKDH----SEPPNKNFKADEEPVE----ETPAEEPVQVER-----E 725
Db      1107 SGIESDDHCQRBQEOLEKENLKLSSENIELKFQLEQANDLPRLKKNQVNDLKEMCEFLKE 1166
Yy      726 KVEAOLKEAV-----LLAKYVDSLSKAAETLAGLRNNLTQTIN- 766
Db      1167 KAERYRKIGHRWGSRSGRTPELEKTIIGLMKVYEKYQOREN-EQLKASGITISEKMA 1224
Yy      767 ---DNNSIMEAEKLLALL 782
Db      1225 NIEQENEKALKELERIKAHLL 1244

RESULT 11
YOH8_YEAST
AC      008236;                                STANDARD;             PRT;   1176 AA.
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Hypothetical 131.4 kDa protein in REX4-ATP19 intergenic region.
       YOLO78W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
       Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX      NCBI_TaxID=4932;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=97321807; PubMed=9178509;
RA      Tzeremia M., Katsoulou C., Alexandraki D.;
RT      "Sequence analysis of a 33.2 kb segment from the left arm of yeast
RT      chromosome XV reveals eight known genes and ten new open reading
RT      frames including homologues of ABC transporters, inositol
RT      phosphatases and human expressed sequence tags.";
RL      Yeast 13:583-589(1997).
CC      -|- SIMILARITY: BELONGS TO THE SINI FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outpost.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
DR      EMBL; Z74820; CAA99089.1; -.
DR      EMBL; Z74819; CAA99088.1; -.
DR      SGD; S0005438; YOLO78W.
TM      Hypothetical protein.

```

SEQ	SEQUENCE	1176 AA;	131378 MM;	ED2830D014E5652 CRC64;
Query Match		3.6%;	Score 150.5;	DB 1; Length 1176;
Best Local Similarity		18.6%;	Pred. No. 1.2;	
Matches 157;		Conservative 118;	Mismatches 254;	Indels 313; Gaps 41.
QY	104	GVIVKVDGKYYVYVLEKDAADNVRRKEELINRQKQEHSHQREGGTPRNDCAVA---	LARSO	160
DB	329	GKIFLTDNK-----NDGQKSDSLMANKGI-----	HGDDSSASGNSVSRDGLTETE	374
QY	161	GRTYDDGGYIFNAPSDIE---DTGDAYIYPHGDHYHYPKNETLSASELAAAFLSG---		214
DB	375	SNNISMESEYINKEKLDLDLNPFTVYSNT-----	KKTVSDLGHESTINDGTAV	421
QY	215	---RGNLSNRTYRQRNSDN-TSFHTNVPVSYNPGTNTNTSNNSTNSQAS-----		262
DB	422	NMRDSKDSNSNEFNARDRITPSSYSGKSLIGSEYSEERSYNNDSSTMEGEMLDS		481
QY	263	--QANDIDBLKQLKLPISQR---IHESGL---VDPQOITSRTARCAVAPPHGDY		312
DB	482	DMQOTNTPS---HSIPSMQKYGIGHDDDDSTLNWYVDKAVLTMNSSR-----		526
QY	313	HTIPYSQMSLEBRIARIIPLYRSNHVWPDSPRPOPSQPRPEPSQPQAPANKLIDSN		372
DB	527	-----HFKERRTVYISGKEPTSLTSSNKKFFSVS		554
QY	373	SSLV---SOLVKKVGEVYVEEKGISRYVFAKDLPSETVKNLESKLQK---QESVSH-----		423
DB	555	SNLTSTRSPDLKGHGR-----TSSTASSSEHKAPKVSQSVLHRARKS		596
QY	424	TITAKKENVAP-----RDOEFYDKAVNLTLEAKA-----LFAKNGRNSPQAL		467
DB	597	TTLTKODHQPVPSPSVHKSKEGNILIEKTDYLVSKPKASQLSNMFKKKKRTNTNSV		656
QY	468	DKL-----LE-----RLNDEST---NKEKLVLDLAFAPITPR		498
DB	657	DVLEHFFPYCGDKVPRYSEMGLEIYIQASKKKYKRNSTFTTKYKRSSTIEFVIGALFLYST		716
QY	499	ERLGRKNSQIEYTEDEVRIALQAD-----KYTSDGYIFDEHDIISDEGDAYVTPPHMG		552
DB	717	EK-KPDN---FEEDQLTIVEDISNPNFSLKIVDEDEGEPFEDN-----		754
QY	553	SNWITGK-----DLSIDKEKVAQAQVYTEKG---ILPSP-----DAD-VKANPTG		593
DB	755	---FGFLDKRSTIOSISDESEVVLCKYDAEAKSONELETPLPRETGGGLMDASTLPLANS		811
QY	594	DSAAIYNNKVGKEKRPIPLVRLPYMVEHTVAVNGNLIIPHKOHYHNKIPAWDDHTYKAP		653
DB	812	DTTDDGTINDLSFYK---PIIG---NDDDIIDKTKGSKTI-----DVTYVLYTP		851
QY	654	N-----GYLLEDLFATIKYVEHPDERPHSNCGWGNASEHVLGK---KDHSEDPN-----		700
DB	852	NNNPKFNYYTISVLVT-----SHIND-----ILVYCKCKMKMPDNEALKV		892
QY	701	--KNF-----KADEEPEETPAPEPEVPOVEETEKYEAOLKEAEV-----LIA		739
DB	893	LCKNYLILDLNDIVLRIDGINKVELISKDARBLEHEKMPDKLPVLIQSNDLTPLTL		952
QY	740	KYTDSKKNATETLAGLRNNLTLOIMDNGNSIMAEKEL-----LALLKGSNPPSSVSKK		794
DB	953	EPLNLSLKADAGAAVA-----IPENTKYTSKRAKKISTRYKKLGLAKOHSSSSVA		1004
QY	795	IN 796		
DB	1005	VS 1006		
RESULT	12			
PTNL_RAT		STANDARD:	PRT:	1175 AA.
AC	062728;	062732;		
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		



DT 16-OCT-2001 (Rel. 40, last annotation update)  
DE Protein tyrosine phosphatase, non-receptor type 21 (p33.1.3.48)  
GN PRNP21 OR PRP2E.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=9510449; PubMed=7805871;  
RA L'Abbe D., Banville D., Tong Y., Stocco R., Masson S., Ma S.,  
RA Fantus G., Shen S.H.;  
RT Identification of a novel protein tyrosine phosphatase with sequence  
RT homology to the cytoskeletal proteins of the band 4.1 family.";  
RL FEBS Lett. 356:351-356(1994).  
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein  
CC tyrosine + phosphate.  
CC -1- TISSUE SPECIFICITY: PARTICULARLY ABUNDANTLY IN ADRENAL GLANDS.  
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
CC TYROSINE PHOSPHATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U17971; AAA62153.1; -;  
DR EMBL; U18293; AAA62154.1; -;  
DR HSSP; 006124; 25HP.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR000387; Tyr\_Phosphatase.  
DR InterPro; IPR000242; Tyr\_PP.  
DR Pfam; PF00102; Y\_phosphatase; 1.  
DR Pfam; PF00373; Band\_41; 1.  
DR PRINTS; PR00935; BAND41.  
DR PRINTS; PR00700; PRTPHPTASE.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00194; PRPC; 1.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; 1.  
DR PROSITE; PS50057; BAND\_41\_3; 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PP; 1.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
KW Structural protein; Cytoskeleton; Hydrolase; Alternative splicing.  
FT DOMAIN 77 241 BAND 4.1-LIKE  
FT DOMAIN 922 1175 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT\_SITE 1109 1109 BY SIMILARITY.  
FT VAAPSPIC 1 839 MISSING (IN ISOFORM 2E).  
SQ SEQUENCE 1175 AA: 133411 MW: 82684 kDa  
Query Match 3.6%; Score 149.5; DB 1; Length 1175;  
Best Local Similarity 19.4%; Pred. No. 1.3; Mismatches 273; Gaps 30;  
Matches 138; Conservative 87; Indels 213; Gaps 30;  
QY 29 KTEMLPDEVSKRGINAQIVKITDGYVTS-----HGDDHYHNGKVPY----- 75  
DB 202 EAEMLYQVEVRMDGYGSEYPAK-DSGGSISIGACLDGIFVHKKNRPPVVRMHDA 260  
QY 76 -----DAISELLMKDPNRYLKDEDIVNEKGVYIKVDGKYVYVYLKDAHADNVRTKE 130  
DB 261 NMSHNSKSFALANKETTIOFQTEDEMT-----AKYWRCLVARNKFRYLQOC 309  
QY 131 EINKOKGSHSRRGGPRRNGAVALARSO-----GRYTDDGDIYNASDI 177  
DB 310 NLQTOATLNSVRRGSSSR-----MSLRKPPRYAMPPPPOLHYNGHYT-----EPF 355

QY 178 EDTGDATVPHGDHYHYIPKNEISASELAAAEFLSGRGNLSNSRTY--RRONSNTSRT 235  
DB 356 ASSQDNVEVPKNGKNGFYCHSQSLDRTOID-----LSGR--IRNGSVYAHSTNLSLNTQRP 408  
QY 236 MWVPS--VSNPGTTFNTSNNSTNSQASQ-----SNDISLKLQLYK-LPLSGRH 283  
DB 409 YLQSPMSNSNPSIPSDVWRPDIYPSHRHSALIPPSYRPTDYESVMKRLNGWHAHRH 468  
QY 284 VES-----DGLVDPQAQITSRARVAVPHGDHYF-----IPYSO 319  
DB 469 SHSLRNLIIGSSVAYSRPDALVYQPEI--REPHLASPQSAHYPNLNYSFHSQAPY 526  
QY 320 MSELERIARIITPLKRSRHNHWPDSPRE-----QSPQPT-----PEPSPPQ 362  
DB 527 PVRRPPVGVAV-----SVPELTINVOLOAQOYPAPIRMROYVRRPPPPY 573  
QY 363 PA---PNLK-----IDNSSLSVLVKRVGEGYVEEKISRYFAKDLPSFTYKNLESK 414  
DB 574 PANSTPDLRSLHLYISSNPDLITRVHHSVO--TFQEDSLPVAHSLOVSEPLTAARAH 631  
QY 415 LSKQESY-----SHTLTAKKENVAPR-----DOEFYDKAVNLLTEAH 451  
DB 632 LQKRSITELIAGITGFEGLRLKEETMSASADVAPRTSAGSQSSVFSQKVKQESTEEQG 691  
QY 452 KALEFNKGRNSDPQAL--DKLLERLNDESTNKEKLVDDL-----AFLAP 494  
DB 692 SGGSYSHKSLSDATWTLHSSEDEDELDSSREHNVSEPRILTAARSQOQLNYPQASVTP 751  
QY 495 ITHPERLCKPMSQIYTYTDEVRIALQADKYTTSQGYTFDEHDIISDEGDVATYTPMGSH 554  
DB 752 VTGPLHIEPKSHV--TFPERKAKDISPVHLYME-----THQPRRH 790  
QY 555 WIGKSLSDKEKVAQAAYTKKGLPPSPDADVKANPTGDSAAATYNRVK 605  
DB 791 GLTPSMSESLTTSGRIRARDSL-----KKRPVSLSLGKKNTVEG 833  
RESULT 13  
ID RBPL\_PLAVB STANDARD; PRT: 2869 AA.  
AC 000798;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, last sequence update)  
DT 01-OCT-1996 (Rel. 34, last annotation update)  
DE Reticulocyte binding protein 1 precursor.  
OS Plasmodium vivax (strain Belen).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
NCBI\_Taxid=31273;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92315338; PubMed=1617731;  
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;  
RT "A reticulocyte-binding protein complex of Plasmodium vivax  
RT merozoites".  
RL Cell 69:1213-1226(1992).  
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO  
CC HUMAN RETICULOCYTE CELLS.  
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M88097; AAA29743.1; -;  
KW Malaria; Receptor; Signal; Transmembrane.  
FT SIGNAL 1 17 POTENTIAL.  
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.

FT DOMAIN 18 2807 EXTRACELLULAR.  
 FT TRANSMEM 2808 2826 POTENTIAL.  
 FT DOMAIN 2827 2869 CYTOPLASMIC.  
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).  
 SQ SEQUENCE 2869 AA: 330213 MW: B9DBE442205EBCFF CRC64;

Query Match 3.6%; Score 149; DB 1; Length 2869;  
 Best Local Similarity 17.2%; Pred. No. 4.9;  
 Matches 166; Conservative 148; Mismatches 336; Indels 314; Gaps 42;

QY 24 KATQKTEMLTDEVSRCRGIAEQ-----IVIKITD-----QGT---VTSH 62  
 DB 1702 KSEYKNETLVQNEMSR---INVEESLTDIDKKTIDIDNLLKMKQYEEGLQKIKEN 1758  
 QY 63 GCHYHYNGKVPYDAIISSELLMKDN-----YKLDEIVNEVGAVIVYDKGYYY 116  
 DB 1759 AD-----KKSINFELVGSINLALDPSTSIPTKILKEIDMTGDLK-NYGVAMNEIHGEF 1812  
 QY 117 LKD---AAHADN-----VTRKEINRQKQSHQRE-----144  
 DB 1813 TKSYNLIEFHLNADTVSVPFKAQSLRELAKREEHLRREBEAFLINDIKVSESLK 1872  
 QY 145 -----GTPPRDGAVALARSGRTTDD-----GYTFNASDITIEDTG 181  
 DB 1873 LKEMKKVSAEYEGMKRDSYVSQIVQDKTIVDELKTINDISECSSYLVNVSIVYKVK 1932  
 QY 182 DAYIVPHGHHYIPKNEHSASELAAEAFLSGRNLNSRTYRQNSDNTS---RTNV 238  
 DB 1933 ES---KHAYRR--DANSYKESKVTLYANFLSDEAKISSGMEAMKSNFKTDELETF 1987  
 QY 239 PVSNGTNTNTNSNSN-----TNSQASOSNDIDSLKQLYKPLQRHVESDGLV 290  
 DB 1988 SVIANSNELLKIEDOSNOVIOKERESQDLAMDATDIYVVK--LKNEPMLEKKNE 2045  
 QY 291 FDBAQITSTAGVAVPHGDHIFLFSOM---SELEETIARIIPRYSNMHWPDSPRE 347  
 DB 2046 EYVSEKVEALRLSQVESIRCFEFHRLDNLTELELMKKWVTI-YR-----DKKSE 2098  
 QY 348 OPSPOPTPEPSPGPAPAPMLKIDNSISVSQIVRVKGEYVEEKGISRYVFAKDLPSST 407  
 DB 2099 RES-----GLQEMEN-EMNTYSNSTOL-----EIVVSAGSKSDIDIKLENSNE 2143  
 QY 408 VNLESKSKQBSVSHTLTAKKENAVAPRODFYDKAYVNLLETAHKALEYNKGRNSDQAL 467  
 DB 2144 MNISIKIS-----TIDSKYIEMNSTIDELYKLGKN--COAHWISLSITYAN--MKT 2191  
 QY 468 DKLLERLNDSNTEKLVDDLAFLAPITHPERLGRPNQOIEYDEVRIGADLQKTTTS 527  
 DB 2192 SKKLIMINKERTEKCV-----YIKDSSST 2219  
 QY 528 DGYI-----FDEHDIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQVYTK 575  
 DB 2220 DGVEVELKFGYSKLFSSASEIYNADITYSVNPAKHE---KESLNATRLDKELLYLFH 2275  
 QY 576 KQILPPSPADYKANPTG-DSAAATYNRVKGKR-----IPLVRLPYMEVHEV-- 624  
 DB 2276 Q-----NSDISIVEGVQNMALAYDKLNEKEREMDELYRNISFTKLQK--EHSYDVK 2327  
 QY 625 -----KGNLITP-----632  
 DB 2328 PMIELHKGNETNNKNSLLEKEKKLSVNDHMSMEAMIKGLKTYTPESVQINNIYVI 2387  
 QY 633 -----HKDHY-----HNKFAMFDDHT-----YKAPGYTLEDLFATI 665  
 DB 2388 EAEVKTLEIDIDYGNVYIVEEHKKQFSIILDRNALMDLDEIRKKNYML-----2440  
 QY 666 KYVEHPDERPHS-NDGWMASAEHVLGKKDSEDPKNKFADEEVEETPAEP-----717  
 DB 2441 ---MEVNETIIRVNDYIKITNKLQVAKTEYQLEIKQDDMLQNFLLKVSITIEYF 2497  
 QY 718 EYPOVETEVEAQLEAEVLLAKVTD--SLKANATEETLAGLRNNLTLOIMNNSIMAA 775

DB 2498 ENVKKKESILNDYQERLL-KIGBHLDEIKRNVETLTSYEIDQKEMMSKN-LLEKK 2555  
 QY 776 ERL 779  
 DB 2556 SKMM 2559

RESULT 14  
 CUT7\_SCHPO  
 ID CUT7\_SCHPO STANDARD: PRT: 1085 AA.  
 AC P24339;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Kinesin-like protein cut7.  
 GS CUT7 OR SPAC25G10.07C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OX NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91015362; PubMed=2145514;  
 RA Hagen I., Yanagida M.;  
 RT "Novel potential mitotic motor protein encoded by the fission yeast  
 RT cut7 gene."  
 RL Nature 347:563-566(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=2184401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeli C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Stevens S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Grynopre B.,  
 RA Weltjens I., Vantreels E., Rieger M., Schefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Gaillet A., Cadieu E., Dreano S., Gloux S., Lelaire V., Motlier S.,  
 RA Lucet M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerruti L., Lowe T., McCombie W.R., Paulsen I., Potshkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RL "The genome sequence of Schizosaccharomyces pombe."  
 RT Nature 415:871-880(2002).  
 CC -!- FUNCTION: COULD BE A SPINDLE POLE BODY MOTOR. ON TRANSITION FROM  
 CC G2 TO M PHASE OF THE CELL CYCLE, THE SPINDLE POLE BODY DUPLICATES;  
 CC THE DAUGHTER POLE BODIES SEED MICROTUBULES WHICH INTERDIGITATE TO  
 CC FORM A SHORT SPINDLE THAT ELONGATES TO SPAN THE NUCLEUS AT  
 CC METAPHASE. MUTATIONS AT CUT7 BLOCK SPINDLE FORMATION.  
 CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BINC  
 CC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL: X57513; CAA40738.1; -
DR EMBL: Z70691; CAA9436.1; -
DR PIR: S14032; S14032.
DR HSP: P17119; 3KAR.
DR InterPro: IPR001752; kinesin_motor.
DR Pfam: PF00225; kinesin_1.
DR PRINTS: PR00380; KINESINHEAVY.
DR SMART: SM00129; KISC; 1.
DR PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE: PS50067; KINESIN_MOTOR_DOMAIN2; 1.
DR Motor protein; Cell division; Microtubules; ATP-binding; Coiled coil;
KW Mitosis; Cell cycle; Phosphorylation; Repeat.
FT DOMAIN 70 435 KINESIN-MOTOR (BY SIMILARITY).
FT DOMAIN 436 604 COILED COIL (POTENTIAL).
FT DOMAIN 715 740 COILED COIL (POTENTIAL).
FT DOMAIN 897 955 COILED COIL (POTENTIAL).
FT NP_BIND 159 166 ATP (BY SIMILARITY).
FT REPEAT 987 998
FT REPEAT 999 1010
FT MOD_RES 1011 1011
FT PHOSPHORYLATION (BY CDC2) (BY
FT SIMILARITY).
FT CONFLICT 34 61 SASNPKRRRPPITDGYPRSDTNSPT ->
FT LRAILDGVDLSLLFL (IN REF. 1).
SQ SEQUENCE 1085 AA; 122133 MW; 5669277875559D58 CRC64;

Query Match 3.5%; Score 145.5; DB 1; Length 1085;
Best local Similarity 18.6%; Pred. No. 2;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATQRTENLTPEVSKREGINAE-----QIVIKI--TDQGVYVSHGDHYHYN 70
DB 159 GGTGTGTYTMSGD-LSDSDGILSEAGLIPRALYOLFSSLDNSNGEYAVK-CEYELYN 216
QY 71 GKAVYDAIISSELLMKRPNTKLKDED-----YNEVKGGVYIKVDKYYIYIKAAADN 125
DB 217 EEL-RDLVSEEL--RKPARVEEDTSRGNVYIGIEESY-IKNAQGLRLIRGSHRQ 272
QY 126 VRTKEEINPKQKHS-----QHREGTFRNDGAVALARSGRYTDDGYFNASDIHEDT 180
DB 273 VAAKCKNDLSRSISITITLHRYVSSGMDENSLITNN-----NSDDLK-- 319
QY 181 GDAYIVPHGHYHYIPNELSASELAEEFLSGRNLMSNRT----- 223
DB 320 -----ASKLHMVDLASENI--GRSGENKRAKETGMINQSLTLGRVI 361
QY 224 -----YR-----RQNSDNTSRTNWPVSINPCTTNTNNSNNTQSASQS 264
DB 362 NALVEKAHHIPRESKLTRELLQDSLGGKRTYSMTIVVS---STVTNLEETSTLEVAARA 418
QY 265 NDIDSLKQLYKLPESQHVESDGLVDFPAOI-----TSRTARGVAVPHGDHYHFIFYSQ 319
DB 419 K--STNNKQNNQNVLRKYLINDVLIDIERKMDLNATRKKNQVYLAESTYKELMDRVQ 475
QY 320 MSEL-----EERIAIIPLRYRSNHWPDSPREDSPOPTPEPSPQAPAPLKI----- 369
DB 476 NKDLCEQARKLEVLDLNYS-----SRQQLQYVSKSNQEHKEVEALQQLVNSSTE 529
QY 370 -----DSSNSLVQVAKVGEVYFEKGISSRYFAKDLSEYTKNLESKSKQESVHT 424
DB 530 LESVYSENEKIKNELVLEIKRKKYETNEAKITTVADLSQYRESKEYTASLVEKIDRT 589
QY 425 LTAKENAVAPDQEFYKAVNLT-----TEAH 451
DB 590 ERNNKEN-----ENNFWMLKKNLTLMLRSFHSFTDETNGYFTLLNDNASKNEELLNTHSN 645
QY 452 KALFYKNGRNSDQALDKLER-----LNDESTNKEKLVDDLAFIAPIT 496
DB 646 QLLISMRTITEHFQSDLEALQASRSCAVPNSSLDLYSELKDKSKNLDLALHSLQDIS 705
QY 497 -HBERLCKRPS-----QIETTEDP-----VA--IAQL 520
DB 706 MSSOKLNGISSELIELQDKMESYROLVQELRSLVNIQHTHESQKELMWGYRNDIDL 765

```

```

QY 521 ADKYYTS--DGYPDEHDILSDGDAVYTPMGHSHWIGKSLSDKEVAAQAYTKEGI 578
DB 766 VKTCTTSLANDDII-LSDYISDQSKFESKQODLIANGIKVSNFLQDONSXYTK----- 820
QY 579 LPPSPDADVKANPTGDSAAAI--YNRVKGKRIPLVR-----LPYVETHEVVK 625
DB 821 -----ADILSHLNDINNSIRKANETIMNNRSEEFLLRAAASQAEIVGANKERIOKTVE-- 872
QY 626 NGNLIIFKHKH-YINIKFAMFDDHTTKAPNG-----YLEDLFATIKYVEHPDERP 676
DB 873 NGSQLDLSKSKRAHSNRSKSMYDHCIALAESQKGVNLEVOYTLDRLOVK-----E 923
QY 677 HSNQWGNASAEHVLGKKDSEDPPKNEKADEPVEETPAPEPEVQVETEKVEAOLEAEV 736
DB 924 HSEDNTEKHQQLLDLESVLGNNDNL-----ISIKPHELO----- 962
QY 737 LLAVYDSSLK-----ANATETLAGLRN---NETLOIMNNSIIMAE-----EKLLA 780
DB 963 ---KITDHLVKGTSTSLANHTELLGLGDESICNLETTEDTSLVKLETTYGDTPSKRELPA 1019
QY 781 LKGSNPSVSKEKIN 796
DB 1020 TPSTWRDSSLIKETTN 1035

```

```

RESULT 15
YNDI_YEAST
ID YNDI_YEAST STANDARD; PRT; 1240 AA.
AC P53935;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 141.5 kDa protein in YPT53-RHO2 intergenic region.
GN YNL091W OR N2231.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C / FY1679;
RX MEDLINE=96367601; PubMed=8771715;
RA Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;
RT *sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
RT chromosome XIV that includes the YPT53, TRN14U and GSR m2 genes and
RT four new open reading frames.*;
RL Yeast 12:599-608(1996).
CC -!- SIMILARITY: TO S.POMBE SPAC29E6.10C.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X65811; CAA59826.1; -
DR EMBL: Z71367; CAA95967.1; -
DR SGD: S0005035; YNL091W.
KW Hypothetical protein.
FT DOMAIN 756 761
SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D65822D5778 CRC64;

```

```

Query Match 3.5%; Score 145.5; DB 1; Length 1240;
Best local Similarity 19.8%; Pred. No. 2.4;
Matches 148; Conservative 85; Mismatches 236; Indels 277; Gaps 35;

QY 220 NSRTYRQNSDNTSRTNWPVSINPCTTNTNNSNTNSQASQNDIDSL-----KQLY 275
DB 4 NSKSKRRKN-----KSKQHNKKKNGNSDPEQSIINPTQLVPRMEPELY 44

```

```
OY 276 KLPISQRHVESDGLVDPQAITSRTARGVAVPHGDHYHPIRYSQMSLEERTARI-- 331
      | | | : : : | : : : | : : :
Db 45 -----HTESD---YPTSRVYIKRAPNGDVIYE-----PINTDDKKERTAN/LTHNKD 87
      | | | : : : | : : : | : : :
OY 332 -----PLRYR-SNHVWVDSRPEQSPQTPPEPSPQAPAPNLKIDSNSLSQLYRK-- 382
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 88 SMDASASSLAFTLDSHW-----ESLSPE-----EKKTLIRLEKEE--VENVIARNYQ 130
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 383 -----VGEYVFEKGISR-----YFAKDLPSET-----VKNLE-SKLSKO 418
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 131 DHSCSCSCVCGRRHLAMQOEMERITNTLYAMDKDPEINPIKPHLIGITIKELQISKNOQ 190
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 419 ESVSHF-----LTKKENVAPRDOEFYDKAYN--LLTEAHKALF 455
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 191 NDLSTKGEVYVNFSLSSTVGSLKEEVLHFQKQKLSQOAHNETADNTSLLEENLNNIH 250
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 456 XNKGK---NSDFQALDKLLERLNDSTNKEK-----LYDD 487
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 251 INKTSSEISANFNVSVD--EELQOKYSNFTKTFISSHPKIAEYVOKMMYPNIRALTD 308
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 488 LL-----AFLAPITHPERLG-----KPNSQLRYTEDEVRIAQLAD-----KYTT 526
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 309 LMNSNGGFLNAIEDFVNDGOIQASKKDDST--TEDEASSTDLTPKRETTMLHSGKPLT 366
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 527 SDGYTFDEHDIISDEGDAYVT-----PHMGSHWIGKDSLSDKEKVAQAAYTK 574
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 367 EDEYADLQRNIAERMNTNAYDTASKKFKDVSQLKELEFTRFMGSRDKKSFRELIIOSFKK 426
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 575 EKGILPPSPDA-----DYKANPTGDSAAATY-----NRVGEKRI 609
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 427 FDGELGFSVLAATLSSCFSSQSKDTSLDTSIYEDEDEBDYDYSYAEADSEEVSEYEG 486
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 610 PLVRLPYWVEHTVEVKNGNLLIPKDHYNLIKFAWFDDHTYKAPNGYTLLEDFATIKYYV 669
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 487 EAVEKP---EHD-EKSNIGRETJHLSY-----DHDHKKQN----- 517
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 670 EHPDERPHSNDGWNASEHVJGKDH-----SEDPNKNFKADEEPVEETPAPEVPOV- 722
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 518 -HPHHYHST---STHSEDELSEEEYISDIELPHDPKHHFRDDDILDGDEDEPEEEDEN 573
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 723 -----ETKEVEAQLKEAEVLLAKVTDSSILKANETELAGLRNNLTQIMDN 768
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 574 EGDDEEDTYDSGLDETDRLEBGRKLIQIATIKLOSRTIMASTYHEKOA-----DN 622
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
OY 769 NSIMAEAEKLLALLKGSNPSYSYSEK 794
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 623 NRL-----KLQLELEEKRRKKREKEE 643
```

Search completed: November 13, 2002, 04:08:32  
Job time : 40 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: November 12, 2002, 15:27:09 ; Search time 83 Seconds  
(without alignments)  
1976.066 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKNRRVSY.....KLIALKGNPSVSKREKIN 796

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: SPREMBL\_21:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhcc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvivirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4163	100.0	802	16	Q97Q08	Q97Q08 streptococc
2	4163	100.0	816	2	Q9AHT9	Q9AHT9 streptococc
3	3236	77.7	844	2	Q9AG74	Q9AG74 streptococc
4	2797.5	67.2	819	16	Q97Q09	Q97Q09 streptococc
5	2786.5	66.9	819	2	Q9ANY3	Q9ANY3 streptococc
6	2772	66.6	839	16	Q9ANY2	Q9ANY2 streptococc
7	1246	29.9	1039	16	Q9ANY1	Q9ANY1 streptococc
8	949	22.8	825	16	Q99XV4	Q99XV4 streptococc
9	944	22.7	825	2	Q93GT5	Q93GT5 streptococc
10	934	22.4	822	2	Q9ZHG7	Q9ZHG7 streptococc
11	679.5	16.3	289	2	Q9AE21	Q9AE21 streptococc
12	223	5.4	792	16	Q99Z76	Q99Z76 streptococc
13	197.5	4.7	1078	5	Q963T1	Q963T1 plasmodium
14	192.5	4.6	1390	5	Q77033	Q77033 dictyostel1
15	178	4.3	565	5	Q15754	Q15754 dictyostel1
16	176.5	4.2	1271	5	Q25860	Q25860 plasmodium

17	175.5	4.2	1236	5	Q9GTX2	Q9GTX2 plasmodium
18	172.5	4.1	5458	5	Q9U459	Q9U459 plasmodium
19	172	4.1	1795	2	Q9LCJ9	Q9LCJ9 staphylococ
20	172	4.1	2478	2	Q9LCH2	Q9LCH2 staphylococ
21	172	4.1	2481	16	Q99GR6	Q99GR6 staphylococ
22	170	4.1	2478	2	Q9RL69	Q9RL69 staphylococ
23	168.5	4.0	1043	10	Q82345	Q82345 arbidopsi
24	166	4.0	1826	5	Q97255	Q97255 arbidopsi
25	166	4.0	3263	5	Q917U3	Q917U3 arbidopsi
26	166	4.0	16215	5	Q9NFS3	Q9NFS3 drosophila
27	165	4.0	891	16	Q99TD3	Q99TD3 staphylococ
28	164	3.9	891	16	Q99TD3	Q99TD3 staphylococ
29	162.5	3.9	2647	5	Q9U4X0	Q9U4X0 staphylococ
30	162	3.9	2703	5	Q9VPL9	Q9VPL9 drosophila
31	162	3.9	2910	10	Q9END5	Q9END5 arbidopsi
32	162	3.9	5322	5	Q9NT64	Q9NT64 drosophila
33	159.5	3.8	1129	16	Q8XLL1	Q8XLL1 clostridiu
34	159	3.8	1810	5	Q20456	Q20456 caenorhabd1
35	158	3.8	1650	5	Q77328	Q77328 plasmodium
36	156.5	3.8	1495	10	Q9LZA8	Q9LZA8 arbidopsi
37	156	3.7	519	10	Q9SIC8	Q9SIC8 arbidopsi
38	155.5	3.7	1038	3	Q9P619	Q9P619 schizosarc
39	155	3.7	869	4	Q9NIF8	Q9NIF8 homo sapien
40	155	3.7	920	4	Q14673	Q14673 homo sapien
41	155	3.7	1734	5	Q9UOL7	Q9UOL7 plasmodium
42	154.5	3.7	1338	5	Q77306	Q77306 plasmodium
43	154.5	3.7	1463	2	Q86919	Q86919 staphylococ
44	154.5	3.7	1516	5	Q96154	Q96154 plasmodium
45	154	3.7	749	5	Q02602	Q02602 plasmodium

## ALIGNMENTS

## RESULT 1

Q97Q08 PRELIMINARY; PRT; 802 AA.  
ID Q97Q08  
AC Q97Q08;  
DT 01-OCT-2001 (TREMUREL. 18, Created)  
DT 01-OCT-2001 (TREMUREL. 18, Last sequence update)  
DT 01-DEC-2001 (TREMUREL. 19, Last annotation update)  
DE Conserved domain protein.  
GN Sp1175.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1313;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=TiGR4;  
RX MEDLINE=21357209; PubMed=11463916;  
RA Tettelein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
RA Peterson S., Heidelberg J., DeBoy R.T., Halt D.H., Dodson R.J.,  
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Rauden D.,  
RA Holtzapfel E., Khouiri H., Wolf A.M., Uitterlank T.R., Hansen C.L.,  
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,  
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
RT "Complete genome sequence of a virulent isolate of Streptococcus  
RT pneumoniae.";  
RL Science 293:498-506(2001).  
DR EMBL: AF007418; AAK75284.1; -  
DR TIGR; Sp1175; -  
KW Complete proteome.  
SQ SEQUENCE 802 AA; 90080 MW; 4E5CB8364EEA1833 CRC64;

Query Match 100.0%; Score 4163; DB 16; Length 802;  
Best Local Similarity 99.9%; Pred. No. 2e-222;  
Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 SYELGLYQARTVKNRRVSYIDGQARTQKTNLTPDEVSKREGINAEDIVIKIRIDGGVYT 60  
|||||

```

Db 7 SYELGLQARYKENVKNNRVSYIDGKQATOKTENLTPEDEVSKREGINAEOIVIKITDOGYVT 66
Oy 61 SHGDHYHYNGKVPYDAIISEELLKMDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 120
Db 67 SHGDHYHYNGKVPYDAIISEELLKMDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 126
Oy 121 AHADVNRKKEEINRQKQESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 180
Db 127 AHADVNRKKEEINRQKQESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 186
Oy 181 GDAYIVPHGDHYHTPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 240
Db 187 GDAYIVPHGDHYHTPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 246
Oy 241 VSNPGTTNTNTSNNSNTNSQASQSDNDISLKLQLYKPLSGRHVESDGLVFPDPAQITSRT 300
Db 247 VSNPGTTNTNTSNNSNTNSQASQSDNDISLKLQLYKPLSGRHVESDGLVFPDPAQITSRT 306
Oy 301 ARGVAVPHGDHYHTPYSSOMSELEERIAIIPLRYSRSHHWVDSRPEOPSPQPTPEPSPG 360
Db 307 ARGVAVPHGDHYHTPYSSOMSELEERIAIIPLRYSRSHHWVDSRPEOPSPQPTPEPSPG 366
Oy 361 POPAENLKIDSNSLSVQLVRKVGEGYVEEKGISRYFAKDLPSETYKNLESKLSKQES 420
Db 367 POPAENLKIDSNSLSVQLVRKVGEGYVEEKGISRYFAKDLPSETYKNLESKLSKQES 426
Oy 421 VSHITLAKKENVAPRDOEYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 480
Db 427 VSHITLAKKENVAPRDOEYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 486
Oy 481 KEKLVDDLAFAPITTHPERLGPNSQIETTEDEVRIAOADKYTTSDDGYIFDEHDIISD 540
Db 487 KEKLVDDLAFAPITTHPERLGPNSQIETTEDEVRIAOADKYTTSDDGYIFDEHDIISD 546
Oy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATKEKGLPPSPDADVKANPTGDSAAAIY 600
Db 547 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATKEKGLPPSPDADVKANPTGDSAAAIY 606
Oy 601 NRKGEKRIPLVRLPYMEHTVEYKNGNLIIPKHDIYHNIFKAWPDDHTYKAPNGYTTLED 660
Db 607 NRKGEKRIPLVRLPYMEHTVEYKNGNLIIPKHDIYHNIFKAWPDDHTYKAPNGYTTLED 666
Oy 661 LEATIKTYVEHDERPHSDGNGNASEHVLGKDKHSEDPNFKADEEPVEETPAEPEVP 720
Db 667 LEATIKTYVEHDERPHSDGNGNASEHVLGKDKHSEDPNFKADEEPVEETPAEPEVP 726
Oy 721 QVETKEVQAOLKEAEVLLAKYTDSSILKANATEETLAGLRNNLTLOIMDNNSIIMAEAEKLLA 780
Db 727 QVETKEVQAOLKEAEVLLAKYTDSSILKANATEETLAGLRNNLTLOIMDNNSIIMAEAEKLLA 786
Oy 781 LKGSNPSVSKEKIN 796
Db 787 LKGSNPSVSKEKIN 802

```

RESULT 2

```

O9AHT9 PRELIMINARY: PRT: 816 AA.
AC O9AHT9:
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Pneumococcal histidine triad A protein.
GN PHTA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_taxid=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N4.
RX MEDLINE=21116976; Pubmed=11179332;
Wizemann T.M., Heinlrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,

```

```

RA Choi G.H., Barash S.C., Rosen C.A., Masure H.R., Tuomanen E.,
RA Gayle A., Brewah Y.A., Walsh W., Barren P., Iathigra R., Hanson M.,
RA Langermann S., Johnson S., Koenig S.;
RT "Use of a whole Genome Approach To identify Vaccine Molecules
RT Affording Protection against Streptococcus pneumoniae Infection.";
RL Infect. Immun. 69:1593-1598(2001).
DR EMBL; AF291695; AAK19155.1; -
SQ SEQUENCE 816 AA; 91519 MW; 5359126A61D27ED CRC64;

```

```

Query Match          100.0%; Score 4163; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 2,1e-222;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 1 SYELGLQARYKENVKNNRVSYIDGKQATOKTENLTPEDEVSKREGINAEOIVIKITDOGYVT 60
Db 21 SYELGLQARYKENVKNNRVSYIDGKQATOKTENLTPEDEVSKREGINAEOIVIKITDOGYVT 80
Oy 61 SHGDHYHYNGKVPYDAIISEELLKMDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 120
Db 81 SHGDHYHYNGKVPYDAIISEELLKMDPNYKDKDEDIVNEVGVIKVDGKYVYLKDA 140
Oy 121 AHADVNRKKEEINRQKQESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 180
Db 141 AHADVNRKKEEINRQKQESHQREGGTPRNDGAVALARSOGRYTTDDGYIFNASDIIEDT 200
Oy 181 GDAYIVPHGDHYHTPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 240
Db 201 GDAYIVPHGDHYHTPKNELSASELAAEAFLSGRGNLSNSTYTRQNSDNTSRINWVPS 260
Oy 241 VSNPGTTNTNTSNNSNTNSQASQSDNDISLKLQLYKPLSGRHVESDGLVFPDPAQITSRT 300
Db 261 VSNPGTTNTNTSNNSNTNSQASQSDNDISLKLQLYKPLSGRHVESDGLVFPDPAQITSRT 320
Oy 301 ARGVAVPHGDHYHTPYSSOMSELEERIAIIPLRYSRSHHWVDSRPEOPSPQPTPEPSPG 360
Db 321 ARGVAVPHGDHYHTPYSSOMSELEERIAIIPLRYSRSHHWVDSRPEOPSPQPTPEPSPG 380
Oy 361 POPAENLKIDSNSLSVQLVRKVGEGYVEEKGISRYFAKDLPSETYKNLESKLSKQES 420
Db 381 POPAENLKIDSNSLSVQLVRKVGEGYVEEKGISRYFAKDLPSETYKNLESKLSKQES 440
Oy 421 VSHITLAKKENVAPRDOEYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 480
Db 441 VSHITLAKKENVAPRDOEYDKAYNLTFAHKALFXNKGNSDFQALDKLERLNDSTN 500
Oy 481 KEKLVDDLAFAPITTHPERLGPNSQIETTEDEVRIAOADKYTTSDDGYIFDEHDIISD 540
Db 501 KEKLVDDLAFAPITTHPERLGPNSQIETTEDEVRIAOADKYTTSDDGYIFDEHDIISD 560
Oy 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATKEKGLPPSPDADVKANPTGDSAAAIY 600
Db 561 EGDAYVTPHMGSHWIGKDSLSDEKVAQAATKEKGLPPSPDADVKANPTGDSAAAIY 620
Oy 601 NRKGEKRIPLVRLPYMEHTVEYKNGNLIIPKHDIYHNIFKAWPDDHTYKAPNGYTTLED 660
Db 621 NRKGEKRIPLVRLPYMEHTVEYKNGNLIIPKHDIYHNIFKAWPDDHTYKAPNGYTTLED 680
Oy 661 LEATIKTYVEHDERPHSDGNGNASEHVLGKDKHSEDPNFKADEEPVEETPAEPEVP 720
Db 681 LEATIKTYVEHDERPHSDGNGNASEHVLGKDKHSEDPNFKADEEPVEETPAEPEVP 740
Oy 721 QVETKEVQAOLKEAEVLLAKYTDSSILKANATEETLAGLRNNLTLOIMDNNSIIMAEAEKLLA 780
Db 741 QVETKEVQAOLKEAEVLLAKYTDSSILKANATEETLAGLRNNLTLOIMDNNSIIMAEAEKLLA 800
Oy 781 LKGSNPSVSKEKIN 796
Db 801 LKGSNPSVSKEKIN 816

```

RESULT 3

```

O9AG74 PRELIMINARY: PRT: 844 AA.
ID O9AG74

```

AC 09AG74;  
 DT 01-JUN-2001 (TREMblrel. 17, Created)  
 DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE phA.  
 GN phA.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21246685; PubMed=11349048;  
 RA Zhang Y., Masi A.W., Barniak V., Mountzouras K., Hostetter M.K.,  
 RA Green B.A.;  
 RT "Recombinant phA protein, a Unique Histidine Motif-containing Protein  
 RT from Streptococcus pneumoniae, Protects Mice against Intranasal  
 RT Pneumococcal Challenge."  
 RL Infect. Immun. 69:3827-3836(2001).  
 DR EMBL: AF340221; AAK26629.1; -  
 SQ SEQUENCE 844 AA; 94769 MW; D738A55290FF8902 CRC64;

Query Match 77.7%; Score 3236; DB 2; Length 844;  
 Best Local Similarity 75.3%; Pred. No. 4.6e-171;  
 Matches 620; Conservative 66; Mismatches 99; Indels 38; Gaps 4;

QY 1 SYELGYAARTVKKNNRYSYIDGKQATQKTENLPDEVSKREGINAEOIVIKITDOGYVT 60  
 DB .21 SYELGYAARTVKKNNRYSYIDGKQATQKTENLPDEVSKREGINAEOIVIKITDOGYVT 80  
 QY 61 SHGHHYHYNGKVPYDAIISELLMKDPNYKLKDEDIYNEVGKGVIVYDGGKYYVYLKDA 120  
 DB 81 SHGHHYHYNGKVPYDAIISELLMKDPNYQLKDEDIISELKGIVIKVDGKYYVYLKDA 140  
 QY 121 AAADNVRTKEEINKOKESHQHSQREGGTPRNDGAVALAASQGRYTTDDGIYFNASDIIDT 180  
 DB 141 AAADNVRTKEEINKOKESHQHSQREGGTPRNDGAVALAASQGRYTTDDGIYFNASDIIDT 200  
 QY 181 GDATVYPRGHGHHYIPKNELSASELAEAFLSGRGNLSNRTYRROSDMSTRNWVPS 240  
 DB 201 GDATVYPRGHGHHYIPKNELSASELAEAFLSGRGNLSNRTYRROSDMSTRNWVPS 260  
 QY 241 VSNPGETTNTSNNSTNSQASQSNDSIDSLKOLYKLPLOSRIHVESDGLVDPDAQITRT 300  
 DB 261 VSNPGETTNTSNNSTNSQASQSNDSIDSLKOLYKLPLOSRIHVESDGLVDPDAQITRT 320  
 QY 301 ARGVAVPRGHGHHYIPYQMSLEERIKARIIPLYRSNHWVPDSRPDPSPQPPPEPSG 360  
 DB 321 ARGVAVPRGHGHHYIPYQMSLEERIKARIIPLYRSNHWVPDSRPDPSPQPPPEPSG 380  
 QY 361 POPAPNLK-IDSN--SSLYSOLVRKVGEGYVFEKGISRYVFAKDLPSERVYKNLESKLS 416  
 DB 381 POPAPNPAPASNPIDELKYEAVRKVGDGYVEFENGVPRIIPAKDLSAETRAAGIDSKLA 440  
 QY 417 KOESVSHTLAKKENVADRDQEFYDRAVNNLLTEAHKALFXNKNRNSDQALDLKLERLND 476  
 DB 441 KOESVSHTLAKKTDLPSSDREFYKAYDILLARIHODLLDNKGRQVDEALDNLLERLKD 500  
 QY 477 ESTNKEKIVYDOLLAFIATIRPERLGRKNSQIETEDEVYRAQLADKTTSDGTYFDEHD 536  
 DB 501 VSSDKVYKLVYDILLAFIATIRPERLGRKNSQIETEDEVYRAQLADKTTSDGTYFDEHD 560  
 QY 537 IISDEGDAYVTPHMGHSWIKDLSDEKVAQAAYTKREGILPSPADVAKANTGDSA 596  
 DB 561 IISDEGDAYVTPHMGHSWIKDLSDEKVAQAAYTKREGILPSPADVAKANTGDSA 620  
 QY 597 AAIYNRVKGERRIPLVRLPYVEHTVEYKNGNLLIPKRDHYHNIFKAFEDDHTYKAPGY 656  
 DB 621 AAIYNRVKAARKVPLDRPYPNLYQTVVEYKNGSLIIPYDHYHNIFKAFEDDHTYKAPGY 680  
 QY 657 TLEDLFAIKTYVEHPDRPHSNDGNGVNASEHYLCKHSDPKNFKADE-----PV 710  
 DB 681 TLEDLFAIKTYVEHPDRPHSNDGNGVNASEHYLCKHSDPKNFKADE-----PV 740

QY 711 EET-----PADEPVOVETEKVEAOULKEAEVILAKVT 742  
 DB 741 EETPREEKQOSKPSPKPTPEPEEESPESESEEEFOVEIKENKREKREHDLGKIQ 800  
 QY 743 DSSLKANATETLAGIRNNLTQIMDNNSIMAEKILALLKGS 785  
 DB 801 NPIIKSNKRETLTGKNNLFTQDNNTIMAEKILALLKES 843

## RESULT 4

QY 097OM9 PRELIMINARY; PRT; 819 AA.  
 AC 097OM9;  
 DT 01-OCT-2001 (TREMblrel. 18, Created)  
 DT 01-OCT-2001 (TREMblrel. 18, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE Conserved domain protein.  
 GN Sp1174.  
 OS Streptococcus pneumoniae.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
 OC Streptococcaceae; Streptococcus.  
 OX NCBI\_TaxID=1313;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TIGR4;  
 RX MEDLINE=21357209; PubMed=11463916;  
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,  
 RA Peterson S., Heidelberg J., Deboy R.T., Hatt D.H., Dodson R.J.,  
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,  
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,  
 RA Holtzman E., Khouli H., Wolf A.M., Uterback T.R., Hansen C.L.,  
 RA McDonald L.A., Feldblyum T.V., Angiolini S., Dickinson T., Hickey E.K.,  
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,  
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;  
 RT "Complete genome sequence of a virulent isolate of Streptococcus  
 RT pneumoniae."  
 RL Science 293:498-506(2001).  
 DR EMBL: AE007418; AAK75283.1; -  
 DR TIGR: SP1174; -  
 SQ Complete genome.  
 KW Complete proteome.  
 QY SFUENCE 819 AA; 92228 MW; 43852B72E8163BDE CRC64;

Query Match 67.2%; Score 2797.5; DB 16; Length 819;  
 Best Local Similarity 67.0%; Pred. No. 8.4e-147;  
 Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;

QY 1 SYELGYA-RTVKKNNRYSYIDGKQATQKTENLPDEVSKREGINAEOIVIKITDOGYV 59  
 DB 21 SYELGYAQQADCKESNNVAYIDGQACQAKENLPDEVSKREGINAEOIVIKITDOGYV 80  
 QY 60 TSHGHHYHYNGKVPYDAIISELLMKDPNYKLKDEDIYNEVGKGVIVYDGGKYYVYLKDA 119  
 DB 81 TSHGHHYHYNGKVPYDAIISELLMKDPNYQLKDEDIISELKGIVIKVDGKYYVYLKDA 140  
 QY 120 AAADNVRTKEEINKOKESHQHSQREGGTPRNDGAVALAASQGRYTTDDGIYFNASDIID 179  
 DB 141 AAADNVRTKEEINKOKESHQHSQREGGTPRNDGAVALAASQGRYTTDDGIYFNASDIID 197  
 QY 180 TGDATVPRGHGHHYIPKNELSASELAEAFLSGRGNLSNRTYRROSDMSTRNWV 239  
 DB 198 TGDATVPRGHGHHYIPKNELSASELAEAFLSGRGNLSNRTYRROSDMSTRNWV 246  
 QY 240 SVSNPGETTNTSNNSTNSQASQSNDSIDSLKOLYKLPLOSRIHVESDGLVDPDAQITSR 299  
 DB 247 NPAQRLSENHLVTPYHQ--NOGENISSLLRELYAARPLSRHVESDGLVDPDAQITSR 305  
 QY 300 TARGVAVPRGHGHHYIPYQMSLEERIKARIIPLYRSNHWVPDSRPDPSPQPPPEPSP 359  
 DB 306 TARGVAVPRGHGHHYIPYQMSLEERIKARIIPLYRSNHWVPDSRPDPSPQPPPEPSP 365  
 QY 360 GPQAPNLKIDSNSSLYSOLVRKVGEGYVFEKGISRYVFAKDLPSERVYKNLESKLOE 419  
 DB 361 GPQAPNLKIDSNSSLYSOLVRKVGEGYVFEKGISRYVFAKDLPSERVYKNLESKLOE 419





```

RX MEDLINE-21357209; PubMed-11463916;
RA Tetteijn H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn O., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Knouti H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblum T.V., Anglinoli S., Dickison T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AF318955; AAK06760.1; -.
DR EMBL; AE007403; AAK75120.1; -.
DR TIGR; SP1003; -.
KW Signal: Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 66.6%; Score 2772; DB 16; Length 839;
Best Local Similarity 65.0%; Pred. No. 2.2e-145;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNRYSYIDGKATQKTENLTPDEVSKREGINAEOIVIKITDQGYV 59
DB 21 SYELGRHQAGVKKSSNNYSYIDGQAGCAKEMLTPDEVSKREGINAEOIVIKITDQGYV 80
QY : 60 TSHGHYHYNGKVPYDAIISELLMKDPNKKLKDIEDIVNEKGGYVIRKVDGKYVYLKD 119
DB 81 TSHGHYHYNGKVPYDAIISELLMKDPNKKLKDIEDIVNEKGGYVIRKVDGKYVYLKD 140
QY 120 AAHADNVRTKEEINROKQEHSGHGTGPRNDGAVALARSOGRTTDDGYIFNASDIID 179
DB 141 AAHADNVRTKEEIKQKQKQEHSHHGGGS--NDQAVAAARQGRITDDGYIFNASDIID 198
QY 160 TGDAYIVPHGDHYHYIPKNELASAEALAAEAFLSGRGNLSNRTYRONSDNSTRNWPV 239
DB 199 TGDAYIVPHGDHYHYIPKNELASAEALAAEAFVWNC-----KQSSRSSSSSYNA 247
QY 240 SVSNNGTNTNTSNNNSQASQSDNDLSLKQLYKPLPSGRHVESGVLFPDQITSR 299
DB 248 NPAQRLSENHMLVTPTPYHQ--NOGENISLSLRELIAKPLSERHVESDGLIFDPAQITSR 306
QY 300 TARGAIVPHGDHYHYIPYSONSELEERTIARIITPLRYSNHWPDSRPPSPQTPPEPSP 359
DB 307 TARGAIVPHGDHYHYIPYEQNSELEKRIARIITPLRYSNHWPDSRPPSPQTPPEPSP 366
QY 360 GPQAPNLK-IDSN--SLVSQLVARKVGEYVPEEGKISRYYFAKDLPSETVKMLSEKL 415
DB 367 SPQAPNPQAPNSNIDRKYKAEAVKGVGDFEENGVSRTIPAKDLSAETAAAGIDSKL 426
QY 416 SKQESVSTTLAKKENVAPRQDEFYDAKYNLLTEAHKALFNKKGNPSQALDKLERLN 475
DB 427 AKQESLSRLAKKADLDSSREFNKAAYDLARITHODLNNKGQVDFEALDNLLERLK 486
QY 476 DESTKEKLVLDLAFAPITPHERLGPNSQIETXEDVIRIAOIAADKYTSDGYIFPEH 535
DB 487 DVPSKAVLVLDLAFAPITPHERLGPNSQIETXEDVIRIAOIAADKYTSDGYIFPEH 546
QY 536 DIISDEGDAYVTPHNGSHWIGKDSLSDKEKVAQAQYKREGILPSPDPAVKAAPTSDS 595
DB 547 DIISDEGDAYVTPHNTSHWIKKDSLSEARAQAQYKREGILPSPDPAVKAAPTSDS 606
QY 556 AAATYNRKGEKRIPLVLPYVETVEYKNGNLIIPKDHYNIKFAMFDDHYTKAPNG 655
DB 607 AAATYNRKAAKAVPLDMKPYNLQYTVENKNGSLIIPYDHNHFKFEMFEGLEYAEKGG 666
QY 656 YLLEDLFATIKYVYHPDERPHSNDGMSAPHYGK-----KQDSE----- 697
DB 667 YLLEDLFATIKYVYHPDERPHSNDGMSAPHYGK-----KQDSE----- 726
QY 698 -----DPRKNFKADEPVEETPAPEVPOVETEKVEAQLKAEAVIL 738
```

```

DB 727 ESDEKENHAGLNPSADNLKYPSTDEETPEEAEEDTTPDEAEIFQVENSVINAKIADAEALL 786
QY 739 AKVTSLSKANATEFLLAGIRNLLTQINDNNSIMAEAEKLLALKGSNPSV 790
DB 787 EKVTDPISRONAMETLTIKSSLLGLTKDNNNTISAEDVSLALIKESOPAPI 838

RESULT 7
Q9ANY1 PRELIMINARY; PRT; 1039 AA.
AC Q9ANY1;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-OCT-2001 (TREMblrel. 18, Last annotation update)
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein
DE SP1004).
GN PHE OR SP1004.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE-21101045; PubMed=1159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Dormitzer M., Dagan R., Brewhar Y.A., Barren P., Lathigra R.,
RA Langermann S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the Pht family) that are protective against sepsis."
RL Infect. Immun. 69:949-958(2001).
RN (2)
RP SEQUENCE FROM N.A.
RX STRAIN-TIGR4;
RX MEDLINE-21357209; PubMed-11463916;
RA Tetteijn H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn O., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Knouti H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblum T.V., Anglinoli S., Dickison T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
DR EMBL; AF318955; AAK06761.1; -.
DR EMBL; AE007403; AAK75121.1; -.
DR TIGR; SP1004; -.
KW Signal: Hypothetical protein; Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT NON_TER 1039 AA; 114631 MW; 81A563FC806625C4 CRC64;

Query Match 29.9%; Score 1246; DB 16; Length 1039;
Best Local Similarity 40.9%; Pred. No. 9.3e-61;
Matches 296; Conservative 97; Mismatches 180; Indels 150; Gaps 18;

QY 1 SYELGLYQARTVYKENNRYSYIDGKATQKTENLTPDEVSKREGINAEOIVIKITDQGYV 60
DB 22 AYALNQHSQENKQNNRYSYVDSOSSQSENLTPDQVSOKEGIAEOIVIKITDQGYV 81
QY 61 SHGHYHYNGKVPYDAIISELLMKDPNKKLKDIEDIVNEKGGYVIRKVDGKYVYLKDA 120
DB 82 SHGHYHYNGKVPYDALFSELLMKDPNKKLKDIEDIVNEKGGYVIRKVDGKYVYLKDA 141
QY 121 AAHADNVRTKEEINROKQEHSGHGTGPRNDGAVALARSOGRTTDDGYIFNASDIID 180
DB 142 AAHADNVRTKEEINROKQEHSGHGTGPRNDGAVALARSOGRTTDDGYIFNASDIID 197
QY 181 GDAYIVPHGDHYHYIPKNELASAEALAAEAFLSGRGNLSNRTYRONSDNSTRNWPV 240
DB 198 GNAYIVPHGDHYHYIPKDSLASELAAKAHLAKGNMOPQSOLSYSSASTSDN----- 248
```

OY		241	VSNFETTNTNTSNNSTNNSQASQSDNDISLKLQYLKPLSQRIHVESDGLVFPDAQTIST	300
Db		249	-----NTQSANGASTSPKANSENIOSSLIKELYSPSAORSESDGLVPDAKIISTRT	301
OY		301	ARGVAVPGHDIYHPIPTSONSELEERTARIITPLRYSNMHVWPDSPREDSPPQTPPEPSSG	360
Db		302	PNGVAIPPGHDYHFIPISKLSALEEKIARWP-----	333
OY		361	POPAPNLKIDNSSLVLSOLVRKVGEGVFEEEKISRYPFAKDLPEETVANLESKLSQES	420
Db		334	-----ISGTGSTVTGMNK-----PNEVSSLSGSLSNPSS	363
OY		421	VSHLTAKKEVAVPRDOEFTDKAYNLILTEAHKLFKNKGANSDFQALDKLERLNDESTN	480
Db		364	-----LTTSKELSSASDCYIENPK-DIYEETATAYIVRHG--DHFHPIPK-----SMQIG	410
OY		481	KEKLVDDLAFLAATITERPERIGRKNSEIETEDEEVRIAQLADKYTTSDDGYTFDEHDIISD	540
Db		411	OPTLPNNSLAPPSP-SLPINFGTSHKHE-----EDGIFGDANITAE	452
OY		541	EGDAYVTPPHMHSHMICKDSLSDKEKVAQAQYTKEGILPPSPDADVKANPNPTGDSAAAIY	600
Db		453	DESQVVMHGDNHYFPFKKDLTBEQIAAKKHLE-----YKSHNGGLDSLSH	501
OY		601	NR-----VGKGRIPLYRLPYVV--EHTEYVENGNLLI-PKHDIYNHIKEM	644
Db		502	BODPSPNAKEMKLDIKTEELKIEKGIIMQCYGVKKREESTIVNKKEMKALIITYPGHDHHHADP--	558
OY		645	FDDHTYKAPNCTYLEDLFATIKTYVEHPDRPHSNQG---WGNASEHLGKKKDSSEDPN	700
Db		559	IDEH---KPVGIG---HSHSNTELFRPEGCAKKKBGNKYTGEEILTNVNLLMKSTENN	611
OY		701	KNF 703	
Db		612	ONF 614	
<hr/>				
RESULT 8				
O99XV4		099XV4	PRELIMINARY; PRT;	825 AA.
AC		099XV4		
DT		01-JUN-2001	(TREMBLrel. 17, Created)	
DT		01-JUN-2001	(TREMBLrel. 17, Last sequence update)	
DT		01-DEC-2001	(TREMBLrel. 19, last annotation update)	
DE		Hypothetical protein SPY2006.		
GN		SPY2006.		
OC		Streptococcus pyogenes.		
CC		Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OX		NCBI_TaxId=1314;		
RN		[1]		
RP		SEQUENCE FROM N.A.		
RC		STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;		
EX		MEDLINE=21192684; PubMed=11296296;		
RA		Ferrucci J.U., Meschan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,		
RA		Primeaux C., Seate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,		
RA		Quan Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,		
RA		Lian X., Clifton S.W., Roe B.A., McLaughlin R.;		
RT		"Complete genome sequence of an M1 strain of Streptococcus pyogenes."		
RL		Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).		
DR		EMBL; AE006623; AKK34688.1;-		
KW		Hypothetical protein; Complete proteome.		
SQ		SEQUENCE 825 AA; 92649 MW; 57DF50969BD50F4 CRC64;		

	Query Match	22.8%	Score 949;	DB 16;	Length 825;
	Best Local Similarity	29.5%;	Pred. No. 1.9e-44;		
	Matches 277;	Conservative 100;	Mismatches 214;	IndeIs 348;	Gaps 34;
Oy	1 SYELGLYQARTVKEKNRRSYID--GKQATOKTEINTLPDEVSAREGINAEQIVIKITTDGG	57			
Db	:: : :::  ::       :	:			
	22 SYOLKLHHNGSATKNOQLAYIIDDSSKGAKAPRT-NKTMOJIASEBEISAEQIVTKITTDDG	80			
Oy	58 YVTSGHDHYHNYGRVPDALISEELLMDKPYYKLTKDEDIVNEVGKYVIKDKGKIYYVL	117			

Db	81	YVTSIGDHYHFNCKNPDAIISSELLMTPTNRFKQSDVINELIDGYIKVNGVYVL	140
Qy	118	KDAHADVVRKEELNROKOEHSQ-REGG-----TPRNCDAVALARSQRYTTDDGY	169
Db	141	KGGRKRNIRFKQDAEDYAKGTKEAKERLAQVAHLKREFAVAANVEAKRQGRITTTDDGY	200
Qy	170	IFNMSDIIEDTGDAYIVPHGDHYHYIKPNKLSASELAFAEFLS---GRGLNSRSTYRR	226
Db	201	IFSPDIIIDGLDAYLVPHGNHMYIKPKDLSPELSLAQAQVMSQKGRG--ARPSDRP	258
Qy	227	QNSDVTSTNM-VFSVS-NQCTTNTNTSN-----NSNTNSQASQSDN-----IDSL	271
Db	259	TPAPAPGRKAPRIDVYVNPQGGQDPNGGYHAPRPDASQNKHQREDFKCTFKELL	318
Qy	272	KQYKLTPLFSQAHVESDGLVPEPPAITSRTARGVAVPHGDHYH IPIQSOMSELEETIARI	331
Db	319	DQHLHDLKTYHVEDGLIEFPPTOYKSNAGFYVVRPHGDHYHTIPSSQSLPEMELAD--	376
Qy	332	PLRYKSNHWVDSRPEQSPDPPTPEPSGPQAPNLIKIDSNSLSQLVRYKVEGYVEE	391
Db	377	---RYLAGGTEDD-----DQSGD-----	391
Qy	392	KGISRYFAKDLPSBYKANKLESKSOESYSHLLYIAKKENAVAPRODEFTDKAYNLLTAAH	451
Db	392	-----HSKPSDKE-VTHFTFLGHR-----IKAY-----	412
Qy	452	KALFYKNGRNSDPQADLKLERLUNDESTNKEKLVDDLAFIAPITPBERLGQPSQIET	511
Db	413	---GKGLD-----GAP-----	420
Qy	512	EDEVRIADLADKYTTSDQYIFDEHDIISDEGDAYVTPHMGSHMIGKDSLDPKEVAAQA	571
Db	421	-----YPTSDAYVSKESHSYVDKSGVYFAKHGDHFHYIGFELDEQYELDEVAN	468
Qy	572	YTKKGGILPSPDPADYKANPFGSAAAIYRYVGEKRIPLY-----RLPTVY	618
Db	469	VWKANG-----QADELAAALDQEQEKER-PLFTYKVVSRKVTYKDGKVGYYM	513
Qy	619	-----EH-----TYEVKN-----GN	628
Db	514	PRDKQGYVANDQDLQIADAEQELMKKKKKHYRIDYDTGIEPLRADVSSLPMAHGN	573
Qy	629	-----LIIPKDHYNIKFAMFDDHTYKAPNGYTLLEDFTATIKYVYEHDPDERPSND	680
Db	574	ATYDTGSSFVIRPHIDHLYHYYSLL-----TDDQATATIKYVQHEVNR--D	618
Qy	681	GN-----GNASEHYLGKKHSDPNKFKADEEVEETPAE-----PE-----	718
Db	619	IWSKPGHEGSSVIRPNYTPLDKRGAMENMOIHSASEVQKALAEGRFATPDGYIFEDPRV	678
Qy	719	-----VPQVE-----TEKVAQKLEAEVILIAKVTDSLSKANTTEFLA	755
Db	679	LAKETFWKDGSSFIPRADDSSLETTINKSLSQASQWQAQOELIAKKN---AGADYDT--	732
Qy	756	GLRNNLTQIIMDNNSIMAEEAKLIALLLKSGNSPSSVSEK	794
Db	733	-----DKPKKQQAQK--SNENQGPSEAKKEE	757
RESULT 9			
O93GTS			
ID	O93GTS	PRELIMINARY:	PRT: 825 AA.
AC	O93GTS:		
DT	01-DEC-2001 (TRMBLrel, 19, Created)		
DT	01-DEC-2001 (TRMBLrel, 19, Last sequence update)		
DT	01-JUN-2002 (TRMBLrel, 21, Last annotation update)		
DE	Histidine triad protein of group A streptococci.		
GN	HTPA.		
OS	Streptococcus pyogenes.		
OC	Bacteria: Firmicutes; Bacillus/Clostridium group; Lactobacillales;		
OC	Streptococcaceae; Streptococcus.		
OX	NCBI_TaxID=1314;		

[1]  
RN SEQUENCE FROM N.A.  
RP STRAIN-SI-9;  
RC Terao Y., Kawabata S., Hamada S.;  
RA "Characterization of a novel histidine triad protein of group A  
RT streptococci";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB073859; BAB1774.1; -  
SQ SEQUENCE 825 AA; 92623 MW; DE4EC199181DFBF CRC64;

Query Match 22.7%; Score 944; DB 2; Length 825;  
Best Local Similarity 29.4%; Pred. No. 3.5e-44;  
Matches 276; Conservative 101; Mismatches 214; Indels 348; Gaps 34;

QY 1 SYELGLYQARTVKNRRSYID--GKQATQKTEMLTPDEVSKREGINAEOIVKITDQG 57  
DB 22 SYQLGKHHMGSLTKDNQAYLIDDSKGGKAKAPRT-NKTMQJSAEIGISAEOIVVAVITDQG 80  
QY 58 VYTHSGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVGKGVIVYDGRKYVYL 117  
DB 81 VYTHSGDHYHYNGKVPYDAIISELLMTDPYRKQSDVINELIDGVIVKNGVYVYL 140  
QY 118 KDAHADNVRTKEELNROKQEHSH-REGG-----TPRNDGAVALARSGRTYTDG 169  
DB 141 KPGSRKKRNIPTKQJAEQVAKGTKEAKKGLAOVAHLSKEEYAAVNEAKRQGRYTTDDGY 200  
QY 170 IFNADIIEDTGDAYIVPHGDHYHYIPKNELSASELAAAEAFLS--GRGNLSNSRTYR 226  
DB 201 IFSPDIIIDDLGDAYLVHGHHYHITPKKDLSPSELAAAOVWSQKQGRG--ARPSDYRP 258  
QY 227 QNSDNTSRTNW-VPSVS-NPGTNTNTSN-----NSNTSQASQSDND-----IDSL 271  
DB 259 TPAPARRKAPIDPVTPNPGQHQPDNGGYHAPRRPNDAQONKHQDEFKRTFKLL 318  
QY 272 KQLYKLPISQAHVESGVLVFPDAQITSTARGAVAPRGDHYHFIPYSQMSLEERIANIT 331  
DB 319 DQHLRLDKTRYHVEDGLTFEPTQYIKSNAFGYVVPHGHHIIPRSQISPLEMELAD-- 376  
QY 332 PLRYSNHNVDPSPREPQSPPTPEPSGPQAPAPNLKIDSNSLSQLVKRVGEGYVEE 391  
DB 377 --RYLAGQTEDD-----DGGSD----- 391  
QY 392 KGISRYVEAKDLPSETVKNLESKLSQESVSHTLAKKENVAPRODFYDRAVNLITTEAH 451  
DB 392 -----HSKPSDKE-VTHFTLGR-----IKAY----- 412  
QY 452 KALPYNKRNSDFOALDKLERLNDESTNKEKLVDDLAFAPITHPRLGKPSQIEYT 511  
DB 413 -----GKGLD-----GKP----- 420  
QY 512 EDEVRIAGLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDSKEKVAQA 571  
DB 421 -----YTSDAVFSKESISHVDSGVYAKKGDFHHTIGFELQYELDEEVAN 468  
QY 572 YTKREKGLPSPSDADVKANPTGDSAAAIYNNRYKGEKRILPV-----RLPYAV 618  
DB 469 WVKAKG-----QADELAALDQEQGKEK-PLFDYTKVSKRYTKDKGVGVM 513  
QY 619 -----EH-----TVEYKN-----GN 628  
DB 514 PKDGKDYFARDOIDLTOIAFAEOELMKDKKHVRYDIVDTGIEPRLLAVDSSLEPMAGN 573  
QY 629 -----LTIIPKHCHYNNIKFAWFDHDTYKAPNGTYLLEDFATIKKYVVEHDPDERPHND 680  
DB 574 ATYDTGSSFVIRPHDHIHIVPRYSL-----TRDQATITKTYWQHPEVAP--D 618  
QY 661 GW-----GNASEHYLGKRDHSEDPKNKFKADEPEVEETPAE-----PE----- 718  
DB 619 IWSKRGHESSGVIPNVYPLDKRAGMPWQIISAEVQKALAEGRFATPGYIIPDPDV 678  
QY 719 -----VPOYE-----TEVEAQLKEAEVLLAKVYDSSISKANATEPLA 755  
DB 679 LAKETFWKDGSSFSIPRADGSLRTINKSDLSQAEWQAOEILAKKN-----AGATDTF-- 732

QY 756 GLRNNLTQIMDNNSIMAEKRLALKGNSPSSYSKER 794  
DB 733 -----DKRKEKQQAQK-----SNENQDPSSEAKSKE 757

RESULT 10  
Q9ZHG7  
ID Q9ZHG7 PRELIMINARY; PRT; 822 AA.  
AC Q9ZHG7;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
OS Hypothetical 92.4 kDa protein.  
OC Streptococcus agalactiae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Streptococcus.  
OX NCBI\_TaxID=1311;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R268;  
RX MEDLINE=99115568; PubMed=9916102;  
RA Spellerberg B., Rozdzinski E., Martin S., Weber-Heynemann J.,  
RA Schmitzler N., Luetticken R., Podbielski A.;  
RT "Lmb, a protein with similarities to the Irai adhesin family, mediates  
RT attachment of Streptococcus agalactiae to human laminin.";  
RL Infect. Immun. 67:871-878(1999).  
DR EMBL; AF062533; AAD13797.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 822 AA; 92386 MW; 80E4EDF313481F98 CRC64;

Query Match 22.4%; Score 934; DB 2; Length 822;  
Best Local Similarity 29.4%; Pred. No. 1.3e-43;  
Matches 271; Conservative 99; Mismatches 235; Indels 318; Gaps 32;

QY 1 SYELGLYQARTVKNRRSYID--GKQATQKTEMLTPDEVSKREGINAEOIVKITDQG 57  
DB 22 SYQLGKHHMGSLTKDNQAYLIDDSKGGKAKAPRT-NKTMQJSAEIGISAEOIVVAVITDQG 80  
QY 58 VYTHSGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVGKGVIVYDGRKYVYL 117  
DB 81 VYTHSGDHYHYNGKVPYDAIISELLMTDPYRKQSDVINELIDGVIVKNGVYVYL 140  
QY 118 KDAHADNVRTKEELNROKQEHSH-REGG-----TPRNDGAVALARSGRTYTDG 169  
DB 141 KPGSRKKRNIPTKQJAEQVAKGTKEAKKGLAOVAHLSKEEYAAVNEAKRQGRYTTDDGY 200  
QY 170 IFNADIIEDTGDAYIVPHGDHYHYIPKNELSASELAAAEAFLS--GRGNLSNSRTYR 226  
DB 201 IFSPDIIIDDLGDAYLVHGHHYHITPKKDLSPSELAAAOVWSQKQGRG--ARPSDYRP 258  
QY 227 QNSDNTSRTNWVPSVS-NPGTNTNTSN-----NSNTSQASQSDND-----IDSL 272  
DB 259 TPAPARRKAP-IDPVTPNPGQHQPDNGGYHAPRRPNDAQONKHQDEFKRTFKLL 317  
QY 273 QLYKLPISQAHVESGVLVFPDAQITSTARGAVAPRGDHYHFIPYSQMSLEERIANIT 332  
DB 318 DQHLRLDKTRYHVEDGLTFEPTQYIKSNAFGYVVPHGHHIIPRSQISPLEMELAD-- 374  
QY 332 PLRYSNHNVDPSPREPQSPPTPEPSGPQAPAPNLKIDSNSLSQLVKRVGEGYVEE 392  
DB 375 --RYLAG-----QTDNDNS-----GSDH----- 390  
QY 392 KGISRYVEAKDLPSETVKNLESKLSQESVSHTLAKKENVAPRODFYDRAVNLITTEAH 452  
DB 391 -----SKPSDKE-VTHFTLGR-----IKAY----- 410  
QY 453 ALEXNKRNSDFOALDKLERLNDESTNKEKLVDDLAFAPITHPRLGKPSQIEYT 512  
DB 411 -----GKGLD-----GKP----- 418  
QY 513 DEVRIAGLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDSKEKVAQA 572

```
Db 419 -----YDSDAYVFSKESISHVDKSGVTAKHGDFHYIGFLEQYELDEVANW 467
OY 573 TREKG-----ILPPSPDADYKAMPDGSAAA-----LYNR----- 602
Db 468 VAKAGQADELVALLDOEGKEKLEPDTKVKSRKYTKDGAVGYIMPKDGDYIYARQOLD 527
OY 603 -----VKGEKR-----IPVRLPYVEHTEVEVANGNLIIPKHCH 636
Db 528 TQIAFAEDELMLKDKKHRYDYIDVTGIEPRLAVADVSSLPMHAGNATYDGTSSPVIPIHD 587
OY 637 YNINFAWDDHTRYKAPNCTYLEDLFATIKIYYEHDEPHSNDG-----GNASH 688
Db 588 IHVVPYSWL-----TRNQATIKIYVQHEVRP-----DVWSKPEHSEGSVIN 632
OY 689 VLGGKHDSDPNKFKADEPEYETPAE-----PE-----VP 720
Db 633 VTPLDKRAQMPNMQIHSAEVQKALAEGRFAAPDGYIFDPDYLAKETFYVAKDGSFSITP 692
OY 721 OVE-----TEKYAOLKEAEVLLAKVTDLSLKANATETLAGLRNNLTQIMDNNSI 771
Db 693 RADGSSLRTINKSDLSQAEMQOAOELLAKN-----AGDATDT-----DKPBE 735
OY 772 MAEAEKLLALIGSNPSSYSKEK 794
Db 736 KOQADK-----SNEQOQFSKSEE 755
```

## RESULT 11

```
OYAE21 PRELIMINARY: PRT: 289 AA.
ID 09AE21
AC 09AE21
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
DE Hypothetical 32.0 kDa protein (Fragment).
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxId=1311;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5531;
RX MEDLINE=21172873; PubMed=11274116;
RA Gränlund M., Michel F., Norgren M.;
RT "Mutually exclusive distribution of IS1548 and GBS11, an active group
RT I intron identified in human isolates of group b streptococci.";
RL J. Bacteriol. 183:2560-2569(2001).
DR EMBL: AJ290952; CAC35985.1; -.
KW Hypothetical protein.
FT NON_TER 289
SQ SEQUENCE 289 AA; 32043 MW; A15A8588EA8140E4 CRC64;
```

Query Match . 16.3%; Score 679.5; DB 2: Length 289;  
Best Local Similarity 53.8%; Pred. No. 3.6e-30;  
Matches 140; Conservative 40; Mismatches 61; Indels 19; Gaps 8;

```
OY 1 SYELGLYQARTYKNNRNVYID---GKAOTKTENTLPDEVSREGINAEQIYIKITDGG 57
Db 22 SYQLGHNHGLTKDNQIAYIDSKGKAPKT-NKTMQISAEBSISEQIYVKTIDGG 80
OY 58 VYTSHGDIHYNGKVPYDAIISELLMKDPYKDKEDIVNEVGKGYIKVKGKYYVL 117
Db 81 VYTSHGDIHYNGKVPYDAIISELLMTDPNRFKQSDVINELIDGYIKVNGNYYVL 140
OY 118 KQAADADNVRTEELTRQKQEHSH-REGG-----TRNDGAVVALASQGRYTTDDGY 169
Db 141 KGSKRKNIRTKQALAEQYAKGTKEAKEGLAQVALLSEEEVAAVNAEARKQGRYTTDDGY 200
OY 170 INASDIIDDTGDAYIVPGDHYHYIPKNELASSELAAEAFLS---GRNLSNSTRYR 226
Db 201 ITSPDIIIDLDGDAYIVPGNHYHYIPKDKLSPSELAAQATWSQKQNG--ARPSDYRP 258
OY 227 QNSDNTSRTNWVPSVS-NPG 245
```

Db 259 TPAPGRKRAP-IPDYTPNPG 277

## RESULT 12

```
OYAE21 PRELIMINARY: PRT: 792 AA.
ID 09AE21
AC 09AE21
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, last annotation update)
DE Putative internalin A precursor.
GN INLA OR SPY1361.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxId=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najat F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL: AE006574; AAK34188.1; -.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam: PF00560; LRR; 4.
DR SMART: SM00370; LRR; 4.
KW Complete proteome.
SQ SEQUENCE 792 AA; 87458 MW; 9D5E32286485ACE0 CRC64;
```

Query Match . 5.4%; Score 223; DB 16: Length 792;  
Best Local Similarity 19.5%; Pred. No. 0.00028;  
Matches 168; Conservative 116; Mismatches 279; Indels 300; Gaps 42;

```
OY 1 SYELGLYQARTYKNNRNVYIDGKAOTKT-ENLNPDEVSKREG--INAEQIYIKITDGG 57
Db 30 TYPRTKQSRKMTSKIKPIPKSKTKNTKHGVAGVDEPTDGGILTKDSYILSKTDGG 89
OY 58 VYTSHGDIHY-----YNGKVPYDAIISELL-----MKDP-NYKDKD 94
Db 90 IVVDHGHSHFLFYADLDKSPFEYLIPKGSALAKAVAAQRAASQGSKVADPHHHYEFNP 149
OY 95 EDIVNEVGKGYIKVKGKYYVLKDAADNVRTEELTRQKQEHSHQREGGTPRNDGAY 154
Db 150 ADIVADALGYTVRHDHRYILKSS-----LSGQTQAOAKQVATRLPQTSSLV 198
OY 155 ALARSQG-----RYTDDGDIYFNASDIIEDTGDAYIVPHGDHYHYIPKNELASSELAAEA 210
Db 199 STATAGTGLHFPTSDGQFNGGIGVTKDSIIVDHGHLPISFADL----- 248
OY 211 FLSGRGNLSNSTRYKRONSDNSTRTNW--VPSVSNPQTTNTSNNSTNSQASQSDND 268
Db 249 -----RQG-----GMAHVADQYDPAKAKKAPAFTHQTPELSEKEVQ 286
OY 269 SLIKQIYKLPISORHVESDGLVDPAQI---TSRTAGVAVPHGDHYHYIPYQWSELE 324
Db 287 EKLAYL-----AEKLIGIDPSTIKRVETQDGKLGLEYPHHDHAYL---MISDIE 332
OY 325 ERLARTIPLRYSNHWVPSRP-----EQLGIDPSTIKRVETQDGKLGLEYPHHDHAYL---MISDIE 332
Db 333 --TGKDIIP---DPAHIEHARELEKHKKVGMOTLRALGDEVEYILDIVRTHDAPTPPSWE 386
OY 357 PSPGPOP---APNLKIDSNS-----SLVSQLVKRVGEGY----- 387
Db 387 KDPNMKKEWLAIVIKLDGSKRDPLOKRLSLPLNL-ETLIGLGFPIPKDISPLVLFKKLK 445
OY 388 --VFEEKIGISRVFAKDP-----SETVYNLESKLSQESVSHTLTAKEN---VA 433
```

Db 446 OLLMKTGVTIDYRFLDNMFQLEGIDISQNNLKDI--SFLSKYKNL--TLVAADNGIEDIR 502  
 QY 434 PRDOEYDKAYULLTEFAHKAFLFXNKGNSDFQALDKL--LERLUNDES-----TNKE 482  
 Db 503 PLQO-----LPMV-----KFLVLSNNKISDLSPLASLHOLQELHIDNNOITDLSVSKRE 552  
 QY 483 KL--VD-----DLTAFIAP-----ITHPERL--GRPN-----505  
 Db 553 SLTVYDLSSNADVDLATQAPKLETLMVNDYKVSHPDLKKNPNLSLSINRAQOSLEG 612  
 QY 506 -----SOIETEDEVRIQAOLADK-----YTTSDGY--TDEHDIIISDEGA 544  
 Db 613 IEASSVYVVEAEAGNOIKSLVLDKOGSLTFLDVTGNQLTSLGEGVNNFTALDILSVSKNQ 672  
 QY 545 YTPPHMGH-----SHWIGKDSLSD--KEKYAAQAYTKREKGLIPSPDADYKAMP 591  
 Db 673 LTNVNLISKNKTYTINIDISH--NNISLADKLNEQHIFPAIKN--PWAYEGSVGNG 727  
 QY 592 TGDSSAAIYNRVKGKRIPLVRLPYMVEETVEKNGNLIIPKHDYHNIAKFAFPDDHTYK 651  
 Db 728 TAEKRAAMATKAKESAQESASHDYNNHNTYDEEGHA-----HEHRDK-----DDHDIH- 776  
 QY 652 APNGYTLLEDLFATIKYYEHDPDE 674  
 Db 777 -----EHEDE 781

## RESULT 13

Q963T1 PRELIMINARY: PRT: 1078 AA.  
 AC 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
 DE Glutamate-rich protein (Fragment).  
 GN GLUHP.  
 OS Plasmodium reichenowii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=5854;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21313688; PubMed=11420113;  
 RA Theisen M., Thomas A.W., Jepsen S.;  
 RT "Cloning, nucleotide sequencing and analysis of the gene encoding the  
 RT glutamate-rich protein (GLURP) from Plasmodium reichenowii."  
 RL Mol. Biochem. Parasitol. 115:269-273(2001).  
 DR EMBL: AF356828; AAK40236.1; -.  
 GN NON\_TER 1  
 FT NON\_TER 1078  
 SQ SEQUENCE 1078 AA; 123581 MW; 8D388D88B223913C CRC64;

Query Match 4.7%; Score 197.5; DB 5; Length 1078;

Best Local Similarity 19.6%; Pred. No. 0.011;

Matches 164; Conservative 136; Mismatches 347; Indels 191; Gaps 37;

QY 7 YQARTVKKENRVSIDGKQATQKTENLTP--DEVSKREGINAOIVIKITDQYV-- 59  
 Db 372 HAVNVNLOENINNNHOLEQERKNISFEKKNIDSEIILPENVEKEIIVDVSPKHNFHE 431  
 QY 60 -----TSHGDHYHYNGKVPYDAITSEELMKDPYKDKEDIVNEVKGVIKYGKYY 114  
 Db 433 TLEQSTSESEHEAVSEKNAHETVEHEAVSQESNPEDAND--GNVSQSNNEINENEF 489  
 QY 115 VYLKAAH--ADNVRT-----KEEI-----NRQKHSG--HREGGCTPRNDGAAVALARS 159  
 Db 490 VSEKSESEHPANESSLEEAHQEETIVPEQNNQESGKLVNDGEGFEAAHQEETIVEQ 549  
 QY 160 QGRYTTDDGYINASDIEDTGDAYIVPHGDYH--YIPKNELSASFLAAAEFLSGR 215  
 Db 550 NNQEGSEKLVNDGEGFEAAHQEETIVPEQNNQESGKLVNDGEGFEAAHHEMFSEE 609  
 QY 216 GULNSRITYRONSNT-----SRTNWVPSVNPCTTNTNTSNNSTNSQASQSDNDISL 270

Db 610 SN-SELHEHEVESESDPEPEHEHEVSESNPEASENNESSIEBAHQEIESBONDSE 668  
 QY 271 LKQLYKLPSQNHVESDGLVFPDPAQ--ITSRTARGAVANPBGDHYHPIYSQSELEBRAR 329  
 Db 669 LNE-----NELVESEKVSSEPAHEVLVSEKVSSEP--AEHEVLVSEKVSSEPAHEVDS 720  
 QY 330 IIPLYRSNHWPPDRSPEQSPQPPPEPS--PGQAPAPNL-----KIDNSSLVS--QLVR 381  
 Db 721 VSE-KSASE--PSEHEVESVSEQSNNESEKEDGVPSPKPFEEIEKVYQPRIVDLOITE 776  
 QY 382 KVGEIVVEEKGISRYVFAKDLPSFTVKNLESKLSKQESVSHTLPAKKENVAPRDOEYD 441  
 Db 777 -----PNEVFDQPNQPEPPEPSFVKIEKVPSEEN-----KHGSVDPVEVE--- 816  
 QY 442 KAVNLTTEHAKALFANKGNSDFQALDKLELNDESTKKEKLVDDLAFLAPITHPERL 501  
 Db 817 -----KENVSVEVEKQNPOLFEI-----PLKRDDE-I 844  
 QY 502 GKPNQIETEDEVRIQAOLADKYTTSDGYTFDEHDIIISDEGDAYVTPHMGSHWIGKDSL 561  
 Db 845 EKIEPELEF--EDVHTEQDLEHKTYVDEPILVEEL-----PSLEHNEVAHAPEI 892  
 QY 562 SDEKYVAQAAYTKREKGLIPSPDAD--VKANPTGDSAAIYNRVKGKRIPLVRLPYWV 618  
 Db 893 VEIEEYF-----PEPNQNNPEPEINEDDKSAHQEIEVEEELPDEDKNEKV 940  
 QY 619 EH-TVEYKNGNLIIPKHDYHNIAKFAFPDDHTYKAPNGYTLLEDLATATIKYYEHDERPH 677  
 Db 941 EHEIVEVEE--ILPEENNEKV-----EHE-----IVEVEEILPE 973  
 QY 678 SNDGNGNASEH-VLKGKDHSEDPNKNFKADEPVEETPAPEPPOVETEKVBAOLKEAV 736  
 Db 974 ENB--NEKVEHEIVEEELPEEDKNEKVEHEIVEEELPE--ENKNEKVEHEIVEE 1029  
 QY 737 LLAQYDSSLKANATETIAGLRNNLTLOIMDNNSIMAEKLLALLKGSNPSSVSKREK 794  
 Db 1030 ILPEITE-----IEVPSQTNNENIE-----TIKPEKK-----NEFSVYEK 1068

## RESULT 14

O77033 PRELIMINARY: PRT: 1390 AA.  
 AC 077033;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE TRPA.  
 GN Dictyostelium discoideum (Slime mold).  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX2;  
 RA Saito J., Adachi H., Sutoh K.;  
 RT "Dictyostelium TRPA homologous to yeast Sn6 is required for normal  
 RT growth and early development."  
 RL J. Biol. Chem. 0:0-0(1998).  
 DR EMBL: AB009080; BAA3143.1; -.  
 DR InterPro: IPR001440; TPR.  
 DR Pfam: PF00515; TPR. 9.  
 DR SMART: SM00028; TPR. 9.  
 SQ SEQUENCE 1390 AA; 160421 MW; BAAB926656002DE CRC64;

Query Match 4.6%; Score 192.5; DB 5; Length 1390;

Best Local Similarity 18.1%; Pred. No. 0.03;

Matches 147; Conservative 131; Mismatches 291; Indels 245; Gaps 35;

QY 15 NNRVSYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQYTSBGDHYHYNGKVP 74  
 Db 690 NNN 736  
 QY 75 YDAITSEELMKDPYKDKEDIVNEVKGVIKTV-----DGKYYVYLKDAHADNVRT 128

```

Db 737 -----DNVNSKNNNDVLDRRYKGLIEREKTSFNGGR-----:|||:
QY 129 KEINROKOEHSOHRGGRPR-----NDGAVALA-----RSQGRYTDG 168
Db 776 NRD-NDSDRGDRNRGRDRIOETREYNNNNNNNNNNNNNNNNNNNNNN 834
QY 169 YIFNASDIIDTGDAYIV--HGDIHYIIPKNELASASELAFAELSGRGLNSRTYR 225
Db 835 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 885
QY 226 RONSNTSRTNWPVS-----SNPGTTNTSNNSTNSQASQSDND-SLLKOL 274
Db 886 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 941
QY 275 YKLPISORVESDGLVFD-PAQITRT-ARGAVAPGHDIHFIPYSOMSELERIARIIP 332
Db 942 ALSPOSSQIKDRREIILDESDINERSKTRSPSI-----VEAEKRETIYI 988
QY 333 LRYRSHWVPDSRPEQSPQTPPEPSPGOPAPNPKI-DSNLSVLQVRYKVEGYVEE 391
Db 989 DKERS-----PPYIIEKPEDEKQVEKVTDEKESLVEKVDKE-----NEK 1027
QY 392 KGSRYVEAKDLSEYVKNLESLKQESVSHLTAKKENVAPRODEFYDKANLLTEAH 451
Db 1028 ESPSSSSSEKELEKEKEKEKEKEKEKEKEKEKEKEKEKEKEKE-----KEVE 1078
QY 452 KALFXKGNHSDFOALDKLER-----LNDESTNKEKLVDDLAPITHPERLGN 505
Db 1079 KQVEEKK---SVKSEKVEKESTTTTNDDEBEG-----LSETTTTKRD 1123
QY 506 SOIETEDEVRIALQADKYTSDGYFDEHDIISDEGDAYVPHMGSHWIGKDSLDE 565
Db 1124 DSKLPTDEKLISSVPTTAAVEQ-----SRDEKLE 1156
QY 566 KYAAQVTEKGLPPSPADVYKANTGDSAAA-IYNRKGEKRIPLVRLPYVEHTEV 624
Db 1157 MDTKDESEK-----KSTTTTAAASESVKPIDEEKSPPTT-----TTTT 1198
QY 625 KGNGLIIP-HKDIHYNIKPFAMFDDHT-----YKAPNGYTLDFATIKYVEHPD 673
Db 1199 TNYTYEPHKKKESKN---DITTTTTTTTTTKSAKSPNSPTRS-----D 1242
QY 674 E-RPHSNDGWNASHVLGKKDHSED---PNKNPKADEEVEETPA---EPEVPOV 722
Db 1243 EYVEPHQ---DASQEIINKRLKLEDITSPSKRLKPDSPSSATTASTPSPESP-L 1296
QY 723 ETEKYEAQLKEAVALAKYTDSSIKANATETLAG 756
Db 1297 KKENPVGETLSPRIKDKSSSSSSSSSSSSSTNTG 1330

RESULT 15
015754 PRELIMINARY: PRT: 565 AA.
AC 015754: 015700;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE Wima (Fragment).
GN WIMA.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RA Loomis W.F., Iranfar N.;
RL Submitted (Aug-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF020407; AAC16025.1; -
FT NON_TER 565
SQ SEQUENCE 565 AA: 62729 MW: B059B6E08201E1CF CRC64;

```

```

Query Match 4.3%; Score 178; DB 5; Length 565;
Best Local Similarity 20.2%; Pred. No. 0.054;
Matches 85; Conservative 60; Mismatches 158; Indels 118; Gaps 16;

QY 24 KQATQKTEML-TPDESKREGINAEOIVKIKTDQ-----GVVTSGHDIHYHYNCKVP-- 74
Db 118 EQPTQFFSNLIQPNVEY-----TELQSDVIQHOQVIEHONSTIEQHQRHOFSSDSSNI 172
QY 75 --YDAIISEELMKDPNRYKLNKDEDIVNEYKGGYVLYKVDKYYVYLKDAHADNVRTKEEI 132
Db 173 GIYDSLIEHPILVSD-----NYVNSLSDSITVGKDDDETFSPNNNNNDINIDINIDI 224
QY 133 NRQKOEHSOHRGGRTPRNDGAVAL-----ARSGRTYTDGTYIFNASDIIDTGDAYI 185
Db 225 DDNNNTNNNNNSGNQENENLSTSSSEHTPPAOSEPIYVTTTSGSNNNNNNNTI----- 276
QY 186 VPHGDHYIIPKNELASASELAFAELSGRGLNS-----NSRTYRONSNDTSRTNWPVS 241
Db 277 -----INNNE-----NISRDDNSNSNNNNNNNNNNNNNNNNNNNN 310
QY 242 SNPGTTNTSNNSTNSQASQSDNDIDSLKQLYKLPISQRHVESDGLVF----- 291
Db 311 NNNNNNNNNNNNNNNNNNNNGEKEKEISFFS-----SDDVNNSEVLFNSTNSDTNS 363
QY 292 -----DPAQITRTARGAVAPPHGDHYHFIPYSOMSELERIARIIPLRYRSHWVPDSRP 346
Db 364 SANSIETVSVTS---AIAAEKDASSF--FNNISASOE-----LQP 399
QY 347 E-QSPQTPPEPSPGOPAPNPKIDNSLSVLQVRYKVEGYVEEKGISRYVFAKDLPS 405
Db 400 QPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQPQP 457
QY 406 E 406
Db 458 Q 458

```

Search completed: November 13, 2002, 04:10:13  
Job time : 93 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OW protein - protein search, using sw model

Run on: November 13, 2002, 04:08:41 ; Search time 27 Seconds  
(Without alignments)  
867.431 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKEKNRVS.....KLALLKGNPSSVSKKEKIN 796

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	US-08-961-083-56	Sequence 56, Appl
2	2649.5	63.6	763	US-08-961-083-66	Sequence 66, Appl
3	1187.5	28.5	447	US-08-961-083-182	Sequence 182, App
4	163	3.9	3696	US-09-134-001C-5080	Sequence 5080, Ap
5	159	3.8	10182	US-09-134-001C-3159	Sequence 3159, Ap
6	147	3.5	812	US-09-513-783A-6	Sequence 6, Appl
7	146	3.5	783	5231168-2	Patent No. 5231168
8	145.5	3.5	1073	US-09-541-782-6	Sequence 6, Appl
9	145.5	3.5	1073	US-09-723-820-6	Sequence 6, Appl
10	144	3.5	1040	US-08-961-083-118	Sequence 118, App
11	141.5	3.4	2314	US-09-816-703A-2	Sequence 2, Appl
12	141	3.4	2431	US-07-920-281C-2	Sequence 2, Appl
13	141	3.4	2431	US-08-466-277-2	Sequence 2, Appl
14	138	3.3	1503	US-08-976-255-14	Sequence 14, Appl
15	137.5	3.3	1610	US-09-513-783A-22	Sequence 22, Appl
16	136.5	3.3	1183	US-08-447-031A-2	Sequence 2, Appl
17	136	3.3	812	US-09-513-783A-4	Sequence 4, Appl
18	136	3.3	1056	US-09-513-783A-32	Sequence 32, Appl
19	134	3.2	2308	US-08-015-973-1	Sequence 1, Appl
20	134	3.2	2308	US-08-448-164-1	Sequence 1, Appl
21	134	3.2	2308	US-08-081-929-2	Sequence 2, Appl
22	133.5	3.2	1664	US-09-599-652-2	Sequence 2, Appl
23	133.5	3.2	1664	US-08-642-846-2	Sequence 2, Appl
24	133.5	3.2	1664	US-09-264-604-2	Sequence 2, Appl
25	132.5	3.2	1507	5268270-2	Patent No. 5268270
26	132	3.2	1164	US-09-457-708-2	Sequence 2, Appl
27	132	3.2	2032	US-09-071-035-458	Sequence 458, App

28	132	3.2	2032	4	US-09-071-035-462	Sequence 462, App
29	132	3.2	2032	4	US-09-071-035-466	Sequence 466, App
30	131	3.1	1093	5	PCT-US93-03077-1	Sequence 1, Appl
31	130	3.1	2115	4	US-09-324-867-5	Sequence 5, Appl
32	129	3.1	2860	2	US-08-826-267-2	Sequence 2, Appl
33	128.5	3.1	916	1	US-08-188-228-48	Sequence 48, Appl
34	128.5	3.1	916	1	US-08-332-643-42	Sequence 42, Appl
35	128.5	3.1	916	1	US-08-333-638-48	Sequence 48, Appl
36	128.5	3.1	1848	4	US-08-296-791-6	Sequence 6, Appl
37	128.5	3.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
38	128	3.1	783	4	US-09-513-783A-176	Sequence 176, App
39	127.5	3.1	2133	2	US-08-670-707A-37	Sequence 37, Appl
40	127.5	3.1	2133	4	US-09-037-601-37	Sequence 37, Appl
41	127.5	3.1	2133	4	US-09-315-179-37	Sequence 37, Appl
42	126.5	3.0	1181	2	US-08-488-940-2	Sequence 2, Appl
43	126	3.0	688	3	US-09-141-047-8	Sequence 8, Appl
44	125.5	3.0	2133	4	US-09-523-656-30	Sequence 30, Appl
45	125	3.0	868	1	US-07-864-004B-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-08-961-083-56
: Sequence 56, Application US/08961083
: Patent No. 6159469
:
: GENERAL INFORMATION:
: APPLICANT: Choi et. al.
: TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
: NUMBER OF SEQUENCES: 452
: CORRESPONDENCE ADDRESS:
: ADDRESS: Human Genome Sciences, Inc.
: STREET: 9410 Key West Avenue
: CITY: Rockville
: STATE: Maryland
: COUNTRY: USA
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
: OPERATING SYSTEM: MSDOS version 6.2
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/961,083
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Brooks, A. Anders
: REGISTRATION NUMBER: 36,373
: REFERENCE/DOCKET NUMBER: PB340P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 309-8504
: TELEFAX: (301) 309-8512
: INFORMATION FOR SEQ ID NO: 56:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 796 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-961-083-56

Query Match 100.0%; Score 4163; DB 4; Length 796;
Best Local Similarity 100.0%; Pred. No. 1.7e-314;
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 SYELGLYQARTVKEKNRVSIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGVYT 60
Db 1 SYELGLYQARTVKEKNRVSIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDGGVYT 60
```







QY 326 RIARILPLRRENNH-----WPPSRPEQBPQPRP----- 355  
 Db 1583 KEAIAIOLANOKSENEARTAIQNEHSNNNGVAAQAKNSGHEIHELWMPDPAHKKSDAKSIDNTKY 1642  
 QY 356 -EPEGPAPAPNLKIDSSNLSVQL--VAKVEGVFEERKISRVAFLADLPSETVKNLE 412  
 Db 1643 NQOSNTINTTPATDEBKOKALDKLAKADGAYNKVDQAQTNOQVSDAKTEAIDITFINIQ 1702  
 QY 413 SKLSQESVSHLTAKKE-----VAPRDOE-----FYDAKYNLL----- 447  
 Db 1703 ANAVAKPSARVELDSKFEDLKROIATNPATEEKEKODAIQRLNGKRDVEKNLIMODRRDN 1762  
 QY 448 -TEARK-----ALFXNKGNSDFQALDKLLERL-----NDESTNKEKLDVDDL 489  
 Db 1763 EVEGKKNIGLOLEETIHANPTKRSD--ALQELQTKFISQTELLINNNKQATNEK---DEA 1817  
 QY 480 AFLATITPERL-----GKPNQIETEDE--VRIALQ-----ADKTTSDGIYFDEH 535  
 Db 1818 KRLLEISKRTTINQAQTNNQVNAKDNQMNELATITPATITKDTAKTAIDKKAQOQV 1877  
 QY 536 DIISEGDQAYVPHMGHSHWIGKDSLDKEKYAAQAATYKEKGLPSPDADYKANPTGDS 595  
 Db 1878 TIINQNNDA-----TDEKAPAKRLVYKAI-----EAKSNTTND 1913  
 QY 596 AAIATYRVKGEKRIPLVRLPYWEHTVEYKNGNLIIPKHDXHNKFEAFDDHTYKAP-- 653  
 Db 1914 TE--REVGAKTNGLEKINNITQPTQTKTNAKQEIINDKAQOLQIINNTPATEEKEQ 1970  
 QY 654 -----NGYTLEDL-----FATIKYVEHPDERP-----HSDNGGNS 686  
 Db 1971 ATRRVNAGIQAQAIQIINNNAHSTQEVNESKTSINATIKSVQPNVYKPTAINSTLTQEAANO 2030  
 QY 687 EHVLRKKHSEDPNKNF-----KADE--EPEVETPAPEVQVETEKEVA-----Q 730  
 Db 2031 KTLINDGNATDDEKEAKQALVTOKLNEQIQKIHESSTQONQVDNYKAQITAIKLINANA 2090  
 QY 731 LKEAEVL-----LAVTDSLEKAN--AT-----ETLAGRNLL-----TLQIND 767  
 Db 2091 HKROAIIINILNLASKSKSDIRANODATTEEKNRTAIQSIDTPLAQRNNINGANTNALVD 2150  
 QY 768 NNSIMAEEKYLALKGSNPSVSSEKRI 795  
 Db 2151 EN-LEDGKOKIQRIVLSTQTKTQAKADI 2177  
 RESULT 5  
 US-09-134-001C-3159  
 ; Sequence 3159, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GTC-007  
 ; CURRENT FILING DATE: US/09/134, 001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 3159  
 ; LENGTH: 10182  
 ; TYPE: PRN  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-3159

Query Match	3.8%;	Score 159;	DB 4;	Length 10182;
Best Local Similarity	18.8%;	Pred. No. 0.044;		
Matches 189;	Conservative 126;	Indels 320;	Gaps 43;	
QY	13 KENNVSYYIDGQAOAKQKHNLFPPDEVSKREGIN-----AEOIVYKIKINDGQVISH	62		
	:::   ::   :   : ::: :		: :	

Dd	5419	OKSBEELVNSNTRSRVE-----EHUNEAKSLNNAKOLRDVAAETINVK--QSSOYINDS	5473
Oy	63	GDHYHYNGKV-PYDAIISF-----ELMK-----DPYKLDKEDIYNEVKGGYVI	107
Dd	5474	TEHORGDOALOEAENIENIGNPPLTNKSIEOKILOOLTDADONALOGSHLEEAENNAIT	5533
Oy	108	KVDCQYVYLLKDAH---AANVATKEIINQOKEHSQHHEGGTPRND-----GAVALARS	159
Dd	5534	GIN--KITALMDAOROKAIEINVOAQOTIPAVNOQTLDBREINTAMQALRDKYGOONNVHQ	5591
Oy	160	QGRYTTDDGYIFNASDIETEDGDAYVTPHGDHNYHPKNEEL--SASELAAEAFLSREGN	217
Dd	5592	QSNFNEDEQKHNHYDMSVQAGOTIIDKLQD--PIKNKNEIDQALINQINTTQALSGENK	5649
Oy	218	L-----SNSRTRYKQNSDNTSRITNWPVSNSPCTTYNTSNNSTNSQASQSDNDLSLK	272
Dd	5650	LHTDOESTNRQIEGLSSLTNAQINAEKDVLAQNAKTRTDVA-----QKLAAREKESAMS	5703
Oy	273	QLYKLPQSORHVESDGLVFPDPAQITSTARAGVAVPRGDHYHFPYSOMSELEERTARIIP	332
Dd	5704	NL-----RDGI-----ONKEDIKRSAYINADPTKYTAIDQALONAMENIINATP	5747
Oy	333	LRYSNHMVPPDSREOPSPQPTPEPSGPAPAPNLKIDSNSSLVSQLYARKYGEGVFEK	392
Dd	5748	-----NVELNKATIEQALSRYQAOQ--DLD	5771
Oy	393	GISRIVYAKDLPESTVYNLES-----KLSQESVSHTTLAK	428
Dd	5772	GVQOJLAANKOAOATQTVGNLSLNDGQKRELINLLINSANTRTKVOEELNKAFELINHAMAL	5831
Oy	429	KENAAPRD-----OEFDYKAYNLLFPAHKALFYXNGRNSDFQALDKLEER	473
Dd	5832	RNSQVNDQYKQSSNTYINEDQPRQHNTDANVN---EAQATINNNAQPVLDKLAIERLTQT	5888
Oy	474	LN-----DESTNKEKLVDDLAFAPITHPERLQKPNISOIEY-----TEDEVR-IAQ--	519
Dd	5889	VNTKRDALHGAKQKLTQOQAAETGIRGLTSLNEPQNAEVAKVTATATTEDEVIRINIREQAT	5948
Oy	520	-----LADKTTT--SDGXIIFDEHD-----IISDGDAYVYPRHGH	552
Dd	5949	TLDTPAMJGLRKSIKDKDNTNNSKYINEDHDOQOAYDANVANNAQOVIDTQATLS-----	6003
Oy	553	SHWIGKDSLDPKEVAAOATYKKEGILPPSPDADYVAMPPTGSA---AAIYNPVKGEKR	608
Dd	6004	-----SPTINQLANAVTQAKSNL---HSDPTLQHDKXSAQOTIAQLODNLMSAQK	6050
Oy	609	-----IPLVRLPIVVEH-----TVEYKGNLI--IPKCDHYH-	638
Dd	6051	MEDSLIDNESTRTOYOHDLTEAQAOLDGMLGALKESIKDYTNITVSGNFINAEPSSKQAYD	6110
Oy	639	-----NFKFAMPDHDHY--KAPN--GYTLEDDL--	661
Dd	6111	AAVONAOINIINGINOPTINKGANTTATQYVKNTKDLGDGHLHEEKKNANOTINLSNL	6170
Oy	662	-----FATIKYVEHPDERPHSDNGWNASAEHLGKRKH-----SEDP	699
Dd	6171	NNAKQAKEMKLVNSASTLDEYQOONLQDTAQOLDNAGELNQSI--ARKDQYKADSKYLNDE	6229
Oy	700	NKNFKADE-----EPEEETPAPEVPQVETEKVEAOJKEAEVLL--AKYTDSLLKANTE	752
Dd	6230	QIKQNYDDAVOAVRETIINETONPELLKANIDQATOSVQNAEFOALGAEKLDNDKQTSSTE	6289
Oy	753	TLAIGLRNNLTQLOIMDNNSIIMAEAEKLLALLKSGNSPSSYSEKXI	795
Dd	6290	-LDGL-TDILT-----DAQREKIREQINSSNRDIOKOKI	6321

RESULT 6  
US-09-513-783A-6  
; Sequence 6, Application US/09513783A  
; Patent No. 6416959  
; GENERAL INFORMATION:  
; APPLICANT: Giuliano, Kenneth A.

```

; APPLICANT: Kapur, Ravi
; TITLE OF INVENTION: A System for Cell Based Screening
; FILE REFERENCE: 97-022-11
; CURRENT APPLICATION NUMBER: US/09/513,783A
; CURRENT FILING DATE: 2000-02-25
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 812
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:
; OTHER INFORMATION: EYFP-DEAD-MAPKDM construct
US-09-513-783A-6

Query Match
Best Local Similarity 19.8%; Pred. No. 0.0068;
Matches 175; Conservative 103; Mismatches 331; Indels 274; Gaps 39;

Oy 53 ITDQGY-----VTSQDHYHYNGKVPYDAIIESEELMKDP-NYKLKDE-----DIV 98
Db 62 VTFRTYGIQCFARFYRDHKKQHDFFKSAPEGVQERTIFFKDDGNYKTRAEYKFECDTLV 121
Oy 99 NEVK-GGYVIKVDG-----KYYVYLKDAHADNVKTEIKRQKQHS-----Q 141
Db 122 NRIELKGIDFKDGNILHKLLEYNSHNHYIMADKQKNGIKVKNKIHNIEDGSVQLAD 181
Oy 142 HREGTFRNDCAVALARSQGRITTDGTYIFNASDIIEDTG-----ATVPHG 189
Db 182 HYQWTFPIGDDGVLL-----PDNHYLSYQSALSKDPNEKRDMVLLFEVTAAGITLG 233
Oy 190 DHYHYIPKNELSASEL-----AAAEA----- 210
Db 234 MDELKYPDEADSAULSLVDALTEPPPELEGTEIKRDMFAALEFEPYDVIQGVTEVTEFTI 293
Oy 211 -FLSGRNLNSRTRYRONSDNTSRTNWPVSYSNPGTNTNTNSNSNSQASQSDNIDS 269
Db 294 PLLDDEKGTNSESKKKCLD-TSQVEGIPS-SKP-TLLANGDHGMEGNTGASPTD--- 347
Oy 270 LKQLYKLPL--LSQRHVESDGLVPDPAQITSRTARGVAVPRPGDHYHFLPYQMSLEBR 326
Db 348 FLEEVNDYPDYQSSQWMEDEASFCFQPOQVLDTDQ---AEPENEH-----RDDG 393
Oy 327 IARIPLRYRSHHWPRDPRPEQSPQTPPEPSQPPAPNLKIDNSSLSVSQLVKVGEG 386
Db 394 LADLLFVSSGPNMASAFTERDNP-----EDSYGMLPCDSF---ASTAVVSGEWSVGAPN 445
Oy 387 YVEEKGISRYYFAKDLRSETVKNLE-----SKLSKQESVSHTLTAKKENVAPRDQEFYD 441
Db 446 SPCSSECVS-----PEVTIETIQPATELSKAEVSVKQQLPAK----- 484
Oy 442 KAYNLITERKALFYXKNGRNSFOALDKILERLND-----ESTNKRKLVDDLLAFLAPLTH 497
Db 485 -----ALETMAEQTTDVVHSPSDTTPGPDTEALAKDIE- 519
Oy 498 PERLKPNQSUIYETEDEVRIADLADKYTTSQGIPEDEHDIIDEGDAY----- 545
Db 520 -----ETIKPDYILANVTOPSTESDMFLQADMELLTGTEAHHANNIILTEPDE 568
Oy 546 -----VTPEHMSHWIGKDSLSDEKRVAA-----QAYTKREKGIILPPSPADAVK 588
Db 569 SSTKQVAPRMEBEIYIPGNDTSPKTEETTLPIKMDLAPREDVLLKRETELAP-----AGSM 624
Oy 589 ANPTDSDAAIYNRKVKGKRIPLVRLPYMAVEHTVEYKNGNLLIIPKDHYNHKKFAMFDH 648
Db 625 VLSIEIEELIANDVRSAAE-IPVAQ-ETVAVSETEVVLALEVLAP----- 666
Oy 649 TYKAPNGVTLLEDLFAPIKYVEHP--DERPHSNDGMGN-ASPHVYGGKKD-HSEDPKNKRX 704
Db 667 -----SPITTLTKDVTPLPLEAERPLVTDMTPSLETMTGKEKAPPTETNLGMA 716
Oy 705 ADEEPEVETPRAE-----PEVPOVETEKVBAQLKEAEVLLAKVTDSLSKANATETIAG 756
```

```

Db 717 KDMSPLEPSEVTLGKDVYILPETKYAEFFNNV-TPLSEEEV--TSVKDMSPSA---ETAP 770
Oy 757 LRNNLTLO-----IMDNNSIMAEKLLALLKGSNPSVSSEK 794
Db 771 LAKNADLHSGTELLVDNS--MAPASDIALPLETKVATVPIDK 811

RESULT 7
5231168-2
; Patent No. 5231168
; APPLICANT: DIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 2
; LENGTH: 783

Query Match
Best Local Similarity 20.2%; Pred. No. 0.0077;
Matches 141; Conservative 106; Mismatches 278; Indels 174; Gaps 35;

Oy 178 EDTGAYIVPHGDHYHYIPKNELSASELAAEAFISGNGLSNSTRYRRO--NSDNTSRT 235
Db 35 BEGSGSKLV-----DNDEGCFEEAHNHNESF---EVSNSLNERNEPVESK----- 77
Oy 236 MWVPSVSNPGTNTNTNSNSNSQASQSDNIDSLKQLYKLPLSQRHVES--DGLV--- 290
Db 78 ---SVTEPAHEIEEVYSESNPEAPNESSSIEAHQE--ELVPQNDDESGESGLVDNE 131
Oy 291 -----FDPAQITS-----RTARGVAVPHGDHYHYIYSQMSLEEKIARI 331
Db 132 EGDPEEPNHEEPPEPQNDSELSENEIVSEKSVSEP-AEHEIVSEKSVSEPAEHV-ETV 189
Oy 332 PLRYRSHHWPRDPRPEQSPQTPPEPS--PGQPPAPNL---KIDNSSLSVSQLVKVG 384
Db 190 SEKSTSE---PAHEVESSEQSNNEPSEKKDGPVSKPEELEKQDVQPKYD--LQITE 244
Oy 385 EGYVE-----EEKGISRYVFAKDLRSETVKNLESKLSKQESVSHTLTAKK-----EN 431
Db 245 PWFVDSQPPQGPVPSFVKIETKVPSEENKKNASVDPVEYKKEKNGVSEVEKQNSQESVBE 304
Oy 432 VAPRQOETDKAYNLITERHKA-----LPANGRNSDFQALDKILERLND 477
Db 305 IPVNDEDEFEDVHTEQLDLDHKTVDPEIYEVEEIPSELHENEVAHPEIYIEIEVFEPENON 364
Oy 478 STNKEKLVDDLLAFLA-PTTHERLGKPNQSUIYETEDEVRIADLADKYTTSQGIYFDEH 535
Db 365 NEFOIINEDKSAHIQHEIYEVEEILPEDDKNEKYEHEIYEVEEILPEDDKNEKG---QH 420
Oy 536 DIISDEGDAYVTPPHGSHHWIGKDSLSDEKRVAAQAYTKREKGIILPPSPADAVKANPTGDS 595
Db 421 ELVEVE---ELIPE-----DDKNEKVEHEIYEVEEILP--EDKNEKGQHEIYE 463
Oy 596 AAATYNRVKGKRIPLVRLPYMAVEH-TVEYKNGNLLIIPKDHYNHKKFAMFDHYKAPN 654
Db 464 VEIILPEDDKNEK-----VEHEIYEVEE--ILPEDDKNEKG-----QHEIYEVE 503
Oy 655 GYTLLEDLFAPIKYVEHPDE-RPHSNDGMGNAS-----EHVLGKKDHSE----- 697
Db 504 EILPEDDKNEKQHEIYEVEEILPEDDKNEKGQHEIYEVEEILPEDDKNEKGQHEIYEVEI 563
Oy 698 -DPNNFRADEEPEVETPRAEPEVPOVETEKVBAQLKEAEVLLAKVTDSLSKANATETIAG 756
Db 564 LPEDKNEKQHEIYEVE---VEEILPEDDKNEKQHEIYEVEEILPEIYE-----IEVPSQ 614
Oy 757 LRNNLTLOIMDNNSIMAEKLLALLKGSNPSVSSEK 795
Db 615 TNNNENIE---TIKPEEK-----NEFSVEEKA 640
```

```
RESULT 8
US-09-541-782-6
; Sequence 6, Application US/09541782
; Patent No. 6284480
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/541,782
; CURRENT FILING DATE: 2000-04-03
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: prt
; ORGANISM: S.pombe
US-09-541-782-6
```

```
Query Match          3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.014;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;
```

```
QY 23 GQAQTKTENLTPDEVSKREGINAE-----QIVIKI--TDQGYVTSBGDHYHYNN 70
DB 147 GGTGCKTYTMSGD-LSDSGDGILSEAGLIPRALYQLFSSLDNSNQEVAVK-CSYIELYN 204
QY 71 GAVPDAIISSEELMKDPYKDKED-----IVNEVKGGYVTKVDGKYVYLKDAHADN 125
DB 205 EEI-RDLVSEEL--RKPARVEDITSRGNVYITGIEEY-ITKAGDGLRLREGSHRQ 260
QY 126 VETKEINQKQEH-----QHREGTPRNDGAVALAROGQRTYTDGYIFNASDIIDET 180
DB 261 VAATKCNDSLSSHSTFTTLHRKVVSGMTDETNSLTINN-----NSDDLRL-- 307
QY 181 GDAIYVPHGDHYHYIPKNELASASELAAPLSGGNLSNSRT----- 223
DB 308 -----ASKLHWDLGASENI--GRSGAENKRARETMGINOSILTLGRVY 349
QY 224 -----YR-----RQNSDNTSRTMWVPSVSPGTNTNTSNNSTNSQASOS 264
DB 350 NALVEKAHHIPYRESKLRLLQDSLGCKTKTSMIVTVS--STNTLEETISTLEYAARA 406
QY 265 NIDSLKQLYLPLSORHVESDGLVFPDAOI-----TSRTARGVAVPHGDHYHFIPYSO 319
DB 407 K---SIRKRPQNNQVLFKRVYLKDLVLDIERKNDLNATRKNGVYLAESTYKELMDRVQ 463
QY 320 MSEL-----EERTARIITPLKRYSNHWVPDSRPQSPQPTPEPSPGQAPAPNLKI----- 369
DB 464 NMDLLOEQEARKLEYDLNVKS-----SRQLOQYVSKSNQEHKKKVEALQLOLVNASTE 517
QY 370 -----DSNSLYSOLVRYKAGEGYVEEKGISRYVAKDLPSETVKNLESKLSKOESVSH 424
DB 518 LRSVASENKKLNELVLELEKRRKRYETNEAKITVATDLSQYRYRESKEYIASLYEKIDRT 577
QY 425 LPAKKENVAPROEFYDKAYNLL-----TEAH 451
DB 578 ERNNKEN-----ENRPNVNLKFNILITMLRSFHGSFTDETNGYFTLLDNFNASMEELLTNSH 633
QY 452 KALFXNKNNSDFOALDKLLER-----LNDESTNKEKLVDDLLAFLAPIT 496
DB 634 QLLISWTKLTFHQSILDALQSAKSSCAVPSNLSLIVSELDKSKSLDLDALSHQDS 693
QY 497 -HPERLGRKNS-----QIYTTDE-----VR--IAOL 520
DB 694 MSSOKLGNISSELELQDKMESYRQLOVELRSLYLNQHTHEESQKELMYGVRNDIDL 753
QY 521 AKKTYTS--DGYIFEDHDIISDEGDAYYVPHGSHHWIKGKDSLSDKEKAAQAYTEKEGI 578
DB 754 VKTCTTSLNDADIT-LSDYISDQSKFESKQODLIANIGKIVSNFLQECNESLYTK----- 808
```

```
QY 579 LPPSPDADYKANPTGDSAAAI--YNRVKGEKRIPIVR-----LPYWEHTVEVK 625
DB 809 -----ADILSHLNDNTSNIRKANEMNNRSEEFLERNAASQAEIVGANKERLQKTV-- 860
QY 626 NGNLIIPKHDH-YHNIRKFMFEDDHTYKAPNG-----YTLIEDLFAITKYVEHPDERP 676
DB 861 NSQLDSSKSKAISHNSRSMYDHCALAESQOGVNLLEVQTLDRRLQKVK-----E 911
QY 677 HSNDGKNSAEHYLGCKKHSDPNKFKADEPVEETPAPEVPQVEPEKVEAOLKEAV 736
DB 912 HSEDNTKEKHQQLDLESLELVGNNDVL-----IDSIKYPHTLO----- 950
QY 737 LIAKYDSSLK-----ANATETLAGLRN--NLTLQIMDNNSIAEA-----EKLA 780
DB 951 ---KITDHYLKQTLTSLANTNTNELLGIDESLCLNLETTIEDTSLVLYLETTGDTPSKRELPA 1007
QY 781 LKGSNPPSSVSKKIN 796
DB 1008 TPSWTRDSSLIKETTN 1023
```

```
RESULT 9
US-09-723-820-6
; Sequence 6, Application US/09723820
; Patent No. 6468760
; GENERAL INFORMATION:
; APPLICANT: Nislow, Corey
; APPLICANT: Sakowicz, Roman
; APPLICANT: Beraud, Christophe
; TITLE OF INVENTION: Antifungal Assay
; FILE REFERENCE: 1015
; CURRENT APPLICATION NUMBER: US/09/723,820
; CURRENT FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/541,782
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1073
; TYPE: prt
; ORGANISM: S.pombe
US-09-723-820-6
```

```
Query Match          3.5%; Score 145.5; DB 4; Length 1073;
Best Local Similarity 18.6%; Pred. No. 0.014;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;
```

```
QY 23 GQAQTKTENLTPDEVSKREGINAE-----QIVIKI--TDQGYVTSBGDHYHYNN 70
DB 147 GGTGCKTYTMSGD-LSDSGDGILSEAGLIPRALYQLFSSLDNSNQEVAVK-CSYIELYN 204
QY 71 GAVPDAIISSEELMKDPYKDKED-----IVNEVKGGYVTKVDGKYVYLKDAHADN 125
DB 205 EEI-RDLVSEEL--RKPARVEDITSRGNVYITGIEEY-ITKAGDGLRLREGSHRQ 260
QY 126 VETKEINQKQEH-----QHREGTPRNDGAVALAROGQRTYTDGYIFNASDIIDET 180
DB 261 VAATKCNDSLSSHSTFTTLHRKVVSGMTDETNSLTINN-----NSDDLRL-- 307
QY 181 GDAIYVPHGDHYHYIPKNELASASELAAPLSGGNLSNSRT----- 223
DB 308 -----ASKLHWDLGASENI--GRSGAENKRARETMGINOSILTLGRVY 349
QY 224 -----YR-----RQNSDNTSRTMWVPSVSPGTNTNTSNNSTNSQASOS 264
DB 350 NALVEKAHHIPYRESKLRLLQDSLGCKTKTSMIVTVS--STNTLEETISTLEYAARA 406
QY 265 NIDSLKQLYLPLSORHVESDGLVFPDAOI-----TSRTARGVAVPHGDHYHFIPYSO 319
DB 407 K---SIRKRPQNNQVLFKRVYLKDLVLDIERKNDLNATRKNGVYLAESTYKELMDRVQ 463
QY 320 MSEL-----EERTARIITPLKRYSNHWVPDSRPQSPQPTPEPSPGQAPAPNLKI----- 369
```

```

Db 464 NKDLCOBARKLEVLDLNIVKS-----SRQLOVYVSKSNQEHKKEVBALOLQIVNSSTE 517
OY 370 -----DSNSIVSOLVRVYEGGYVEEGKISRYVPAKDLPESTVKNLESKLSKQSVSHT 424
Db 518 LESVASENEKLNELVLEIEKRRKYEETNEAKITTYATDLISOYRRESKEYIASLYEKLRT 577
OY 425 LTAKKENVAPRDOEFYDKAYNLL-----TEAH 451
Db 578 ERNNKEN-----ENFMNKKFNLLTMRSFHGSFTDETNGYFTLLDNFANAMEELLNTSHN 633
OY 452 KALFYKNGRNSPFOALDKLER-----LNDESTNKEKLYVDLLAFAPIT 496
Db 634 QLLISMRTITEFOSILDEALQASRSCAVPNSSLDIYSELKSKNSLIDLAEHSLQDIS 693
OY 497 -HPELKGPN-----QIEYFDE-----VR--IAQL 520
Db 694 MSSOKLNGISSELEILOKMKESYROLVQELRSLYNLOHTHEESOKELMTGVRNDIDAL 753
OY 521 ADKYTTS--DGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAAYTKERGI 578
Db 754 VKTCTSLNDADII--LSOYISDQKSKFESKQODLIANIGKIYSNFLQONESLYTK----- 808
OY 579 LPSPDADYKANPTGDSAAAI--YNRVKGKRIPLVR-----LPYVHEHTVEVK 625
Db 809 -----ADILSHLNDTNSNIRKANEIMNNSSEELFRNAASQAEIVGANKERIOKTV-- 860
OY 626 NGNLIIIPKHD-YHNIRKAMPEDHTYKAPNG-----YTLFEDLFAITIKYVEHPDERP 676
Db 861 NQSOLLDSKRAIHSNRSKMTDHCIALAESQOGVNLVEYTLDRLLQK-----E 911
OY 677 HSNDSWGNASEHVLGKSHSEDPNKNKFADEEPVETPAEPPOVETEKVEAQLKEAEV 736
Db 912 HSEDNTKFKHQQLDLLSIVGNNDL-----IDSKIPHTELQ----- 950
OY 737 LLAQYDSSLK-----ANATETLAGLR-----NLTLQIMDNNSIAEA-----EKLLA 780
Db 951 ---KITDVLKGTSTLANHNTMELLGDESLCNLETTIEDTSIVLFTGTGTPSKRELP 1007
OY 781 LKGSNPSSVSKEKIN 796
Db 1008 TPSTWTRDSSLIKETTIN 1023

RESULT 10
US-08-961-083-118
; Sequence 118, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders

```

```

; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1040 amino acids
; TYPE: amino acid
; STRADEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-118

Query Match 3.5%; Score 144; DB 4; Length 1040;
Best Local Similarity 19.2%; Pred. No. 0.017;
Matches 154; Conservative 111; Mismatches 284; Indels 252; Gaps 38;

OY 107 IKVDGKYV-YLKDAAHADNVTKKEINROKQSHSQHREGGTPRND-----G 152
Db 76 LKIEGYIYGIK-TKQDNTELSRTVDGKYSQSDQPNSTKTSVDVHSADEWNOGQ 134
OY 133 AVALARSQGRITTDGTYFNASDIITEDTGDAYIVPHGDHYHYPKNELSASELAAAEFL 212
Db 135 KVSLL--QGEASGDDG-----LSEKSSIAADMLSSNDSFA 166
OY 213 SGRGNLSNRYRQRNSDNTSRTNWPVSNGTNTNNSNNSNQSASQNDISLKL 272
Db 167 S-----QVEONPDHKGESVRYPTVPEQGNVSATTVOSAEVEVLATND-----R 211
OY 273 QLYKLPLSQRHVESDG-----LVF-----DPAQITS 298
Db 212 PEYKLPLETKGQEPGHBGEAAVREDLPVYTKPLETKGQGBHGEAAVREEPAYTER 271
OY 299 RTARGVAVPHGDHYHIFTYSQMSLEEKRIARIPRYSNMHWNPSPRQPSQPTPPS 358
Db 272 LATKGTQEPGHE-----GRATVEETLEYTEPAVATKGT-----QEBEHGEERYEEL 319
OY 359 PGPQPA-----PNT-----KIDNSNSIVSOLVR--KVGEYVEE 391
Db 320 PALEYTTNRNRETIQNIPTTEIEQDPTLLKNRKIERGQAGTRIQIYEDYIVNGNYET 379
OY 392 KGISRYVPAKDLPESTVKNLESKLSKQSVSHTLTAKKENVAPRDOEFYDKAYNLL--TE 449
Db 380 KEVSRTEVA--PVNEVYKVGTLVKKPTVELTNLTKEYN-----KKSITVSYNIDTTS 431
OY 450 AHKALFYKNGRNSPFOALDKLERLNDESTNKEKLYVDLLAFAPIT--TH--PERLQPN 505
Db 432 AYVS-----AKTQVFHG--DKLVKEVDIENPAKEQVIG--LDYVTPYTVKTHLYTMIGENN 484
OY 506 SQIETDEVRIAOLADKYTTSIDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLSKE 565
Db 485 E--ENTETSTQDFOL--EYKKIEIDIDSELYGKENRY-----RRYL--SLSEAP 530
OY 566 KYAQAAYTKERGIPLPSPDADYKANPTGDSAAAIYNRVKGEKRIPLVRPYVVEET----- 621
Db 531 TDTAKYFVKKV-----SDRKEMLPVPKS-----ITETDGT 563
OY 622 -VEYKNGNLIIPKCHYINIKFAFDHTY-----KAPNGYT--LEDLFAITIK-----Y 667
Db 564 KYTAAVLDOLVERGTDGYK-----DGYTFYAKSKAQPGVYTSFKOLVJAMOSNLGCV 616
OY 668 YVEHDERPHNSDQGNASEHYLGKKDS-----EDPKNKADE--EPVEET 713
Db 617 YTLASDM-----TADVSLGDKQTSYLTGATGSLIGSDGKTSYAIYDLKKPLPFD 667
OY 714 PAEPVPOVETEKVEAQLKEAEVLLAKYTDSSL-----KANATETLAGLRNNLTLQI 765
Db 668 LNCATVRLDLIKTYSADSKENVAALAKANSANINNAVEGIGSAGKAVAGIVASATNTV 727
OY 766 MDNNSIMAEKLLALKGSN 786
Db 728 IENSSFTG---KLIANHODSN 745

```

```
RESULT 11
US-09-816-703A-2
; Sequence 2, Application US/09816703A
; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 AGY
; CURRENT APPLICATION NUMBER: US/09/816,703A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Krueger, N.X. and Saito, H.: A human transmembrane protein-tyrosi
; OTHER INFORMATION: ne-phosphatase, PTP zeta, is expressed in brain and has an N-term
; OTHER INFORMATION: inal receptor domain homologous to carbonic anhydrases; Proc. Nat
; OTHER INFORMATION: l. Acad. Sci. USA 89 (16), 7417-7421 (1992)
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Levy, J.B., et al.; The cloning of a receptor-type protein tyrosi
; OTHER INFORMATION: ne phosphatase expressed in the central nervous system; J. Biol.
; OTHER INFORMATION: Chem. 268 (14), 10573-10581 (1993)
; NAME/KEY: SIGNAL
; LOCATION: (1)..(24)
; OTHER INFORMATION: By similarity
; NAME/KEY: DOMAIN
; LOCATION: (25)..(1635)
; OTHER INFORMATION: Extracellular (potential)
; NAME/KEY: misc.feature
; LOCATION: (25)..(2314)
; OTHER INFORMATION: Mature chain; protein-tyrosine phosphatase zeta
; NAME/KEY: DOMAIN
; LOCATION: (34)..(302)
; OTHER INFORMATION: Carbonic-anhydrase like
; NAME/KEY: SITE
; LOCATION: (105)..(105)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (134)..(134)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (223)..(223)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (232)..(232)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: DOMAIN
; LOCATION: (312)..(406)
; OTHER INFORMATION: Fibronectin Type-III
; NAME/KEY: SITE
; LOCATION: (324)..(324)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (381)..(381)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (497)..(497)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (501)..(501)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (501)..(501)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
```

```
; LOCATION: (552)..(552)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (587)..(587)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (602)..(602)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (629)..(629)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (637)..(637)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (677)..(677)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: VARSPLIC
; LOCATION: (755)..(1614)
; OTHER INFORMATION: Splicing variant; missing (in short isoform)
; NAME/KEY: BINDING
; LOCATION: (997)..(997)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1017)..(1017)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1050)..(1050)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1082)..(1082)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1122)..(1122)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1456)..(1456)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (1548)..(1548)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: BINDING
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1561)..(1561)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1617)..(1617)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: TRANSMEM
; LOCATION: (1636)..(1661)
; OTHER INFORMATION: Transmembrane region; potential
; NAME/KEY: DOMAIN
; LOCATION: (1662)..(2314)
; OTHER INFORMATION: Cytoplasmic (potential)
; NAME/KEY: CONFLICT
; LOCATION: (1722)..(1728)
; OTHER INFORMATION: Missing (in ref. 2)
; NAME/KEY: DOMAIN
; LOCATION: (1744)..(1997)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: ACT_SITE
; LOCATION: (1932)..(1932)
; OTHER INFORMATION: Active site; by similarity
; NAME/KEY: DOMAIN
; LOCATION: (1998)..(2314)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: misc.feature
; LOCATION: (2222)..(2222)
; OTHER INFORMATION: Ancestral active site
US-09-816-703A-2
```

```

Query Match      3.4%, Score 141.5; DB 4; Length 2314;
Best Local Similarity 19.2%; Pred. No. 0.095;
Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 43;

QY 51 IKITOGVTHSGDHYHYNGVPPDAIISELLMKDPYKIKKEDIYNEVGKGYIKVD 110
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 954 VGVTVQGSIFSGPSH-----PIPKSLITPVASLLOPHALSG-----D 993

QY 111 GKYYVYLMDAAADNVRKKEEINRQKQESHQREGTPRNDGAVL-----ARSQGRY 164
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 994 GEM-----SGASSDS-----EFLLPDTDLTALNISSPVSAEFTYT 1030

QY 165 T----DDGVTFNASDIIEDTGATVPHGDHYH-----IPK-----NELSASELAAE 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1031 TSVFDDKALKSEKSELIYGNETELQIPSENEVYSESTVPMNMDVNNKLNASLOETISV 1090

QY 210 AFLSGR-----NLSNRTYR-RQNDNNS-----RTNNVP 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1091 SLSSTKGMFGSLAHTTKVPDHEISOVPENNFSVQPTHVSQAAGDTSLKPLSANSEP 1150

QY 240 SVSNPGTT-----NTNTSNNSNTNSQAS-QSNDISLLKOLYKPLSQRHVES 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1151 ASSDASSEMSPSTQLLFETYSASFTEVLLQPSFQASDVDTLKTIV--LPA---VPS 1204

QY 287 DGLVEDPAQ-----ITSRTARGVAVPHGDHYHFIYQMSSELEERITRIIPRY 335
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1205 DPLVETPKVDKISTTMLHLIYNSGASSEMMLHSTSVPFVDFVSHMNASLOGLTISY 1264

QY 336 RSNHNVPPSRPEQSPQTPPEPSGPQAPAPNLIKIDNSLSLVQLYKKGEGYVEPEKGIS 395
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1265 ASEKEPEVLLKSESSHQVVP-----SLYSNDELFOGTANLEINOAH--PPKG-- 1308

QY 396 RVYFAKDLPS--ETVKNLESKLSKQESVSHLTAKKENAVR-----DQEFYDAX 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1309 RHVEFTPVLSIDELNTLINKLIHSDET---LTSKSSVTGKVFPGITPVASDITVSTDH 1365

QY 445 NL-LTEAR-----KALFXNKG----- 459
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1366 SVPIGNHVAITAVSPHRDGSVTSKLLFPKATSELSHSAKSDAGLVGGEGDPTDDG 1425

QY 460 -----RNSDFOALDKLERLNDSESTNKEKLVDDLLAFLAPTHPERLCKPNSQIY---- 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1426 DDDDDRRDSDGSLIHKCMS-CSSYSRESQEKVMND-----SDTHENSLMDQNNPISLSE 1478

QY 511 -TEDEVRIALQADKXTT-----SDGYIFDENDI-----ISDEG 542
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1479 NSEENRKYTVSSDSQGTGMDNSPGKSPSANGLSQKHNDGK--EENDIQTGSALELSPES 1536

QY 543 DAYVTPHMGHSHMIGK--DSLSDKEKYA--AQAYTKEK--GIL-----PSP 583
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1537 KAMAVLTSDEESGGGSDSLNENETSTDFSFADTNEKDADGILAADSETTPGFPQSP 1596

QY 584 DADV-----KANPTGSAALYINRYKGEK--IPVLRPLVMEHTVEYVKN 627
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1597 TSSVTSENSEVFHVSEAEASNSHESRIGLAEGLESEKKAIVPLIYVSALEFICLVVLG 1656

QY 628 NL-----IIPKHDIYHNKFKFMPEDH--YKAPNGY 656
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1657 ILIYRKCFOTAHFLIEDSTSPRVLTPTPTFPISDVGALPIYKHFKAHADLHASSGF 1716

QY 657 TLEDLFAATK-YVVE-----HPDERPHSN 679
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1717 TEE--FETLKEFYQEVQSCYVDLGITADSSNHPDK--HKN 1753

```

```

TITLE OF INVENTION: Alphaviruses
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESS: Birch, Stewart, Kolasch & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/920,281C
FILING DATE: 13-AUG-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 828-103P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-241-1300
TELEFAX: 703-241-2848
TELEX: 248345
INFORMATION FOR SEQ. ID NO. 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2431 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-920-281C-2

Query Match      3.4%, Score 141; DB 1; Length 2431;
Best Local Similarity 19.7%; Pred. No. 0.11;
Matches 137; Conservative 93; Mismatches 253; Indels 212; Gaps 34;

QY 56 QGYVTHSDHGHYNG-KVPYDAITSEELLKDPYKIKKEDIYNEVGKGYIKVDKRY 114
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1523 KGYSTDSLSLYFEGTGFENQAAIDMAEILTLWPLQANQICLYALGETM----- 1574

QY 115 YLKDAHADNVRKKEEINRQKQESHQREGTPRNDGAVALARSQGYTTDDGYTFNAS 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1575 -----DNIRSKCPVN-----DSDSTPPT--VPLICR-----YAMTAE 1606

QY 175 DIIEDTG-----DAYIVPHGDHYHYIPKNELASELAAEAFLSGRNLSRTYR 225
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1607 RIARLRSHQVSMVYCSFPLP--KYHVDGVQKVCCKVLLFPDTPV--SVSPRYA 1660

QY 226 RQNSNTSRT-----MNVPYSVSNPGTTNTNNTSNNSNTNSQASQSDNDISLLKOLYKPL 279
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1661 ASTYDHSQRSLRGEDLW-----TTDSSSTASDTMSLPSLQSCDIDSIYEPAPIV 1712

QY 280 -SORHVESDGLVFPDPAQTSRTARGVAVPHGDHYHFIYQMSSELEERITRIIPRYSN 338
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1713 TADVHPBPAGIADLAADVHPPA-----DHV-----DIENPILP---PPRPRA 1752

QY 339 HWVDSRPEQSPQTPPEPSGPQAPAPNLIKI-----DSNSSLVSQLYRKVGEGYVEEKG 393
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1753 AYLASRAAERPPVAPR-KPTPAPRTAFRNKLPFGDEDEHEVDALASGITFGEDVDL 1811

QY 394 ISR---YVFAADLSEYVKNLESKLSKQESVSHLTAKKENAVARPDGEFYKATNLLTEA 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1812 LGRAGATIFSSDTGS---GHLOQKSVRQ---HNQCAQDAVOEKKYPPK---LDTER 1861

QY 451 HKALF-----XNKRNSDFOALDKLERLNDSESTNKEKLVDDLLAFLAP---ITHP 498
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1862 EKLLIKMQMHPSEANKSRYS-RKVENMKATVYDRLTSGARLTGADVGATPIYAVYVP 1920

QY 499 ERLGPNQIYETDEVRIA-----OLADKYTTSQY-LFDEHDIISDEGDAYVTPHMGHS 553
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1921 RPVYSPVIERFSSPDVAIACNEVLSRNPYTVASVYQITDEYDAYLDVMD----- 1970

```

```

RESULT 12
US-07-920-281C-2
Sequence 2, Application US/07920281C
Patent No. 5739026
GENERAL INFORMATION:
APPLICANT: Garoff, Henrik
Applicant: Liljestrom, Peter
TITLE OF INVENTION: DNA Expression Systems Based on

```





REFERENCE/DOCKET NUMBER: 229/182  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (213) 489-1600  
TELEFAX: (213) 955-0440  
TELEX: 67-3510  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1503 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
US-08-976-255-14

Query Match 3.3%; Score 138; DB 4; Length 1503;  
Best Local Similarity 19.4%; Pred. No. 0.09; Mismatches 276; Indels 228; Gaps 38;  
Matches 150; Conservative 118;

QY 34 TPDEYSKREGINABQIYIKITDQGYVTSBGD-----HHYNG-KVPYDAIISELL 84  
DB 582 TGPESLQTLARVSELESSTDEDFQSTDPKDSLPGDLHVTSGPESPNNIF----- 636  
QY 85 MKDPYKIKKDDIYNEKKGVIYKVDGKYIYLKDAHADNVKTEINROKQEHSHQRE 144  
DB 637 ---NDVKSSEDLPSSHOKIFDIMEING-VQADFRPATLSSLD-----NPKESYITGHFE 686  
QY 145 GGTFRNDGAVALARSGQYTTDDGYTFNASDI-IEDTGDATVPHGHYHIIPKNELSSAS 203  
DB 687 KEKPKK-----IFDSEPLCLSDN-----LKHQDNF-----DPLWQ 717  
QY 204 ELAAEAELSGRNLNSRTYRONSNDTSRTNWPVSVPNGTT-----NTNTSNNSNTNS 259  
DB 718 ELSEMFLEQKNNLKGSLSKSEHINDQT-----ELKNGFTFAMLETSCRNSLDL 771  
QY 260 QASQNDIDSLKQYIKPLSQRYVESDG-----LVFDPAQITSRRARGAVAPHGDHYH 313  
DB 772 QFAEKPGSL-----QENVSTKGDYDVTGDTLSTLSSPQVQV----- 816  
QY 314 FIPYSQMSLEFRIARIIFLRYRSNHWPDRPSPQPT-----PEP-----SPGP 361  
DB 817 ---PFS--PETETPRRV-----PPDSLPTQGTGPTCLDYIYVEDCLHODISFDA 862  
QY 362 QPAPNLKIDSNSLVSQLVRKVGEGYVEEKGISRYFAKDLPESTYKNLSKLSKQESV 421  
DB 863 VTVR-VEILSTDARTHSIDNR-----SQSPGSEPTL--RLTESDSV 902  
QY 422 -SHLLAKKENVAR---DOEYDKAYNLLTEAHKALFXAKGRNSPQALDK----- 469  
DB 903 LADDILASRVSVGSSLPRLGQELHNKRPSEDSHSHRRLKEMLEAVETLNOJNSKDAKEA 962  
QY 470 -LBERLNDSENNKRIYDLDLAFLAPTHPRGLKPNQSIQYETDEVARIAQLADKYTTSD 528  
DB 963 GLVGLASDSSTQSLSLEDSLSAPRASEPS-LETPDS-LETSV----- 1003  
QY 529 GYIFDEHDIISDEGDVAYTPHMGSHWIGKDSLDEKREAAQAYTK-----KGILPSP 583  
DB 1004 ---DVHEALDLSGSH-TP-----OKLVPRPKPADSGYETVLESPETWLPAP 1048  
QY 584 DADYKANP--TGDSAAAIYNKVGEEKRIPLVRLPYMVEHTYEVANGNLIIPKDHYNIAIK 641  
DB 1049 EGTADSEPATGGDGH-----GLPNPVI-----VISDAGDGRHGEVTP----- 1089  
QY 642 FAWPDDHTYKAPNGYTTLEDLFATIKYUYEHDERPHSNDGMSAENHGLKKHSDENK 701  
DB 1090 -----EFTTGSQGSYRD-----SAFSDNDSEPKR-----SEEPYG-----TSFSA 1127  
QY 702 NFKADEPVEETPAPEYVPOVETEKVEAQLEAF--VLLAKVTDSLSKANPT 751  
DB 1128 LVVQEOPLPEVLPBQSPAQDSCLFARKSQPDESCLSALHNSSDLELRAT 1179

RESULT 15  
US-09-513-783A-22

Sequence 22, Application US/09513783A  
Patent No. 6416959  
GENERAL INFORMATION:  
APPLICANT: Giuliano, Kenneth A.  
APPLICANT: Kapur, Ravi  
TITLE OF INVENTION: A System for Cell Based Screening  
FILE REFERENCE: 97-022-L1  
CURRENT APPLICATION NUMBER: US/09/513,783A  
CURRENT FILING DATE: 2000-02-25  
NUMBER OF SEQ ID NOS: 180  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 22  
LENGTH: 1610  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence:  
US-09-513-783A-22

Query Match 3.3%; Score 137.5; DB 4; Length 1610;  
Best Local Similarity 19.8%; Pred. No. 0.11; Mismatches 337; Indels 265; Gaps 41;  
Matches 174; Conservative 103;

QY 53 ITDQY-----VTSBGDHYHYNGKVPYDAIISELLMKDP-NYKLKDE-----DIY 98  
DB 62 VTTGCGYGLQCFARYPDHNKQIDFFKSNMPEGYVQERTFFQDDGNYKTRAEYKREGDTLY 121  
QY 99 NEVK-GGYVIRKVDG-----KYYVYLKDAHADNVKTEINROKQESH---Q 141  
DB 122 NRIELKIDFKEGDNILGKLEUYNSHNIVYIMDKQKNGIKVNFKIRHNIEDGSQLAD 181  
QY 142 HREGTPNDQAV-----ALARSQ-----RYTTDDGY-----IFNAS 174  
DB 182 HYQNTPTGDBGVLLPDNHYLSYQSALSCKDPNEKRQHWLLEFVTAAGITLMDLEYKKG 241  
QY 175 DIEDTGDVAYIVPHGDHYHYP--KNELASASELAAEA-----PLS 213  
DB 242 DEVGMADLSLV---DALTEPPEIEGEIKRDPMALEAEFYDDIVGTEVTEKTERPILD 298  
QY 214 GRGNSNSRTYRONSNDTSRTNWPVSVPNGTTNTSNNSNTNSQASQSDIDSLKQ 273  
DB 299 GDEKGTGNSKKKPCLD-TSQVEGIPS--SKP-TLLANDHGMENNTAGSPFD--FLEE 352  
QY 274 LYKLP---LSQRHESDGLVDPDAQITRTARGAVANPHGDHYHIFIPYSQMSLEFRIARI 330  
DB 353 RVDYPRDIOSSQNMPEBDASFCFQPOOVLDTDQ---AEPENH-----RDDGLADL 398  
QY 331 IPLRYRSNHWPDRSPPEQSPQTPPEPSPQAPNLKIDSNSLVSQLVRKVGEGYVE 390  
DB 399 LFVSSGPTNASAFTERDNPS-----EDSYGMLPCDSF---ASTAVVSGEMSVGAPNSPCS 450  
QY 391 EKGISRYFAKDLSEYVKNLE-----SKLSQESVSHITLAKKENVAPRQOEYDKAYN 445  
DB 451 ESCVS-----PEVTLETOPATELSKAAEVESVKKQLPAK----- 485  
QY 446 LITEAHKALFXNKGNSFOFOLDKILERLND-----ESNNKKIYDLDLAFLAPTHPRRL 501  
DB 486 -----ALETMADQTTDVVHSPSDTTPRGPTAEALAKLE----- 520  
QY 502 GKPNQIETEYDEYRIQLADKYTTSDGYIFDEHDIISDEGDV----- 545  
DB 521 -----EITRPVILANVTPSTESDMFLAQDMELLGTTEAANNNIILTPEDESSTK 573  
QY 546 -VTPHMGSHWIGKDSLDEKREKVA-----OAYTKEGILPSPDADYKANPT 592  
DB 574 DVAPRMEEEIYPGNDTSPKETEFTLPIKMDLAPPEVLLTKETELAP---AKGWSLS 629  
QY 593 GDSAAAIYNKVGEEKRIPLVRLPYMVEHTYEVANGNLIIPKDHYNIAIKFAMFODHTYKA 652  
DB 630 EIEELAKNDYRSAE-IPVAO-ETVVSETEYVLADEVLP----- 667  
QY 653 PNGYTLDELFAITIKYUYENP--DERPHSNDGMGN-ASEHVLGKKD-HSEDPNKNKFADEE 708

```
Db 668 -----SDITFTLRKDVTLPLEAERPLVTDMPPSLETENTLGKETAAPTETNIGMAKDMS 721
Qy 709 PVEETPAE-----PEVPQVETEKVEAQKAEVYLLAKYTDSSIKANATETLAGLRNN 760
Db 722 PLPESEVITGKDVVILPETKVAEFNNV--TPLSEEV--TSVKDMSPSA--ETEAPLAKN 775
Qy 761 LTLQ-----IMDNNSIMAEAEKLLALLKGSNPSVSKEK 794
Db 776 ADLHSGTELIVDNS--MAPASDLALPLETKVATVPPIKDK 812
```

Search completed: November 13, 2002, 04:13:25  
Job time : 43 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 12, 2002, 15:30:59 ; Search time 47 seconds  
(without alignments)  
255.074 Million cell updates/sec

Title: US-08-961-083-56  
Perfect score: 4165  
Sequence: 1 SYELGLYQARTVKNRVS.....KILALKGNSPSVSKKIN 796

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 97044 seqs, 15060890 residues

Total number of hits satisfying chosen parameters: 97044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCR\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/PCRUS\_PUBCOMB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	4163	100.0	796	US-09-765-272-56	Sequence 56, App1
2	2649.5	63.6	763	US-09-765-272-66	Sequence 66, App1
3	1187.5	28.5	447	US-09-765-272-182	Sequence 182, App
4	170	4.1	2478	US-09-815-242-5816	Sequence 5816, Ap
5	170	4.1	2478	US-09-815-242-12967	Sequence 12967, A
6	168.5	4.0	1043	US-09-946-805-4	Sequence 4, App11
7	159	3.8	870	US-09-815-242-5493	Sequence 5493, Ap
8	159	3.8	870	US-09-815-242-12637	Sequence 12637, A
9	151.5	3.6	2086	US-09-815-242-5639	Sequence 5639, Ap
10	151.5	3.6	5795	US-09-815-242-12610	Sequence 12610, A
11	145.5	3.5	2368	US-09-815-242-5635	Sequence 5635, Ap
12	145.5	3.5	2368	US-09-815-242-12389	Sequence 12389, A
13	144	3.5	1040	US-09-765-272-118	Sequence 118, App
14	143	3.4	2025	US-09-815-242-5703	Sequence 5703, App
15	143	3.4	3158	US-09-815-242-12611	Sequence 12611, A
16	142.5	3.4	6281	US-09-815-242-12996	Sequence 12996, A
17	141.5	3.4	2437	US-09-815-242-5834	Sequence 5834, Ap
18	141	3.4	2437	US-09-901-106-2	Sequence 2, App11
19	140	3.4	1111	US-09-815-242-12955	Sequence 12955, A

20	139.5	3.3	1609	US-09-938-275-11	Sequence 11, App1
21	138.5	3.3	1125	US-09-974-298-114	Sequence 114, App
22	138.5	3.3	1349	US-09-815-242-5898	Sequence 5898, Ap
23	138.5	3.3	1349	US-09-815-242-13137	Sequence 13137, A
24	138.5	3.3	1703	US-09-801-368-340	Sequence 340, App
25	138	3.3	1029	US-09-815-242-5885	Sequence 5885, App
26	138	3.3	1048	US-09-815-242-13083	Sequence 13083, A
27	137	3.3	1016	US-09-815-242-5845	Sequence 5845, Ap
28	135.5	3.3	2076	US-09-815-242-5815	Sequence 5815, Ap
29	135.5	3.3	2165	US-09-815-242-12913	Sequence 12913, A
30	135	3.2	2665	US-09-864-761-34248	Sequence 34248, A
31	134	3.2	1163	US-09-893-348-18	Sequence 18, App1
32	134	3.2	2308	US-10-000-954-2	Sequence 2, App11
33	133.5	3.2	1607	US-09-938-275-10	Sequence 10, App1
34	132	3.2	1164	US-09-950-0464-2	Sequence 2, App11
35	131	3.1	915	US-09-817-514A-6	Sequence 6, App11
36	130.5	3.1	704	US-09-801-368-218	Sequence 218, App
37	130.5	3.1	2434	US-09-815-242-5835	Sequence 5835, Ap
38	128	3.1	1230	US-09-727-384-8	Sequence 8, App11
39	127	3.0	1018	US-09-815-242-5797	Sequence 5797, Ap
40	127	3.0	1018	US-09-815-242-12838	Sequence 12838, A
41	126	3.0	1202	US-09-864-761-43061	Sequence 43061, A
42	124.5	3.0	789	US-09-995-587A-1	Sequence 1, App11
43	124	3.0	1501	US-09-924-154-17	Sequence 17, App1
44	123.5	3.0	1601	US-09-862-027-40	Sequence 40, App1
45	123.5	3.0	1781	US-09-738-877-3	Sequence 3, App11

## ALIGNMENTS

RESULT 1  
US-09-765-272-56  
Sequence 56, Application US/09765272  
Patent No. US20020061545A1

GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8512  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 796 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 56:

US-09-765-272-56

Query Match 100.0%; Score 4163; DB 10; Length 796;  
Best Local Similarity 100.0%; Pred. No. 1.5e+265;  
Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKEENRVSVIDGKQATQKTENTLPDEVSKREGINAEOIVIKITDQGYV 60  
1 SYELGLYQARTVKEENRVSVIDGKQATQKTENTLPDEVSKREGINAEOIVIKITDQGYV 60  
DB 1 SYELGLYQARTVKEENRVSVIDGKQATQKTENTLPDEVSKREGINAEOIVIKITDQGYV 60  
QY 61 SHGDHYYNGKVPYDALISELLMKDPNKLKDEDIVNKGKGYIKVDGKYYVYLKKA 120  
61 SHGDHYYNGKVPYDALISELLMKDPNKLKDEDIVNKGKGYIKVDGKYYVYLKKA 120  
DB 61 SHGDHYYNGKVPYDALISELLMKDPNKLKDEDIVNKGKGYIKVDGKYYVYLKKA 120  
QY 121 AHADNVRTEKEINROKQESHQHEGCTPRNDGAVALARSGRTTDDGYIFNASDIIEPT 180  
121 AHADNVRTEKEINROKQESHQHEGCTPRNDGAVALARSGRTTDDGYIFNASDIIEPT 180  
DB 121 AHADNVRTEKEINROKQESHQHEGCTPRNDGAVALARSGRTTDDGYIFNASDIIEPT 180  
QY 181 GDAYIVPHGDHYHYPKNEISASELAEEAFISGRGNLSNRTYRRONSDNTSRTNWPVS 240  
181 GDAYIVPHGDHYHYPKNEISASELAEEAFISGRGNLSNRTYRRONSDNTSRTNWPVS 240  
DB 181 GDAYIVPHGDHYHYPKNEISASELAEEAFISGRGNLSNRTYRRONSDNTSRTNWPVS 240  
QY 241 VSNPRTNTNTSNNSTNSQASQNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 300  
241 VSNPRTNTNTSNNSTNSQASQNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 300  
DB 241 VSNPRTNTNTSNNSTNSQASQNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 300  
QY 301 ARGVAVPHGDHYHYPYQSELEERLARIPLRYRSNHVPPDSRPEOPSPOTPEPSPG 360  
301 ARGVAVPHGDHYHYPYQSELEERLARIPLRYRSNHVPPDSRPEOPSPOTPEPSPG 360  
DB 301 ARGVAVPHGDHYHYPYQSELEERLARIPLRYRSNHVPPDSRPEOPSPOTPEPSPG 360  
QY 361 POPAPMLKIDSNSLSVQAVRKVGEVVEEKGISRYVFAKDLPSFTVKNLESKLSKQES 420  
361 POPAPMLKIDSNSLSVQAVRKVGEVVEEKGISRYVFAKDLPSFTVKNLESKLSKQES 420  
DB 361 POPAPMLKIDSNSLSVQAVRKVGEVVEEKGISRYVFAKDLPSFTVKNLESKLSKQES 420  
QY 421 VSTHTLAKKENVAPROEFYDKAYNLTFAHKALEFNKGRNSFOALDKLELNDESTN 480  
421 VSTHTLAKKENVAPROEFYDKAYNLTFAHKALEFNKGRNSFOALDKLELNDESTN 480  
DB 421 VSTHTLAKKENVAPROEFYDKAYNLTFAHKALEFNKGRNSFOALDKLELNDESTN 480  
QY 481 KERLVDDLAFIAPITHPERLGRKNSQIEYTEDEVRILAOIADKTTSDGIIFENHIIISD 540  
481 KERLVDDLAFIAPITHPERLGRKNSQIEYTEDEVRILAOIADKTTSDGIIFENHIIISD 540  
DB 481 KERLVDDLAFIAPITHPERLGRKNSQIEYTEDEVRILAOIADKTTSDGIIFENHIIISD 540  
QY 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKERKILPPSPDADYKANPTGDSAAAY 600  
541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKERKILPPSPDADYKANPTGDSAAAY 600  
DB 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAAYTKERKILPPSPDADYKANPTGDSAAAY 600  
QY 601 NRKGGKRIPIVRLPYMVEHTVEYKNGNLIIPKHDYHNKTFAMPDHTYKKAANGYTLTD 660  
601 NRKGGKRIPIVRLPYMVEHTVEYKNGNLIIPKHDYHNKTFAMPDHTYKKAANGYTLTD 660  
DB 601 NRKGGKRIPIVRLPYMVEHTVEYKNGNLIIPKHDYHNKTFAMPDHTYKKAANGYTLTD 660  
QY 661 LEFATIIYYEHPDERHSNDGWNASEHVLGKKHSEDPNNKKADEEPEETPAPEVP 720  
661 LEFATIIYYEHPDERHSNDGWNASEHVLGKKHSEDPNNKKADEEPEETPAPEVP 720  
DB 661 LEFATIIYYEHPDERHSNDGWNASEHVLGKKHSEDPNNKKADEEPEETPAPEVP 720  
QY 721 OVETKEVEAOLKAEVLLAKVTDSISKANATETLAGLRNNLTQIMDNNSIMAEEKLLA 780  
721 OVETKEVEAOLKAEVLLAKVTDSISKANATETLAGLRNNLTQIMDNNSIMAEEKLLA 780  
DB 721 OVETKEVEAOLKAEVLLAKVTDSISKANATETLAGLRNNLTQIMDNNSIMAEEKLLA 780  
QY 781 LKGSNPSVSKEKIN 796  
781 LKGSNPSVSKEKIN 796  
DB 781 LKGSNPSVSKEKIN 796

RESULT 2  
US-09-765-272-66

; Sequence 66, Application US/09765272  
; Patent No. US20020061545A1  
; GENERAL INFORMATION:  
; APPLICANT: Choi et. al.  
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
; NUMBER OF SEQUENCES: 452  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 66:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 763 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 66:  
US-09-765-272-66

Query Match 63.6%; Score 2649.5; DB 10; Length 763;  
Best Local Similarity 66.8%; Pred. No. 2.9e+166;  
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;

QY 1 SYELGLYQARTV-KEENRVSVIDGKQATQKTENTLPDEVSKREGINAEOIVIKITDQGYV 59  
1 SYELGLYQARTV-KEENRVSVIDGKQATQKTENTLPDEVSKREGINAEOIVIKITDQGYV 59  
DB 2 SYELGLYQARTV-KEENRVSVIDGKQATQKTENTLPDEVSKREGINAEOIVIKITDQGYV 61  
QY 60 TSHGDHYHYNGKVPYDALISELLMKDPNKLKDEDIVNKGKGYIKVDGKYYVYLKKA 119  
60 TSHGDHYHYNGKVPYDALISELLMKDPNKLKDEDIVNKGKGYIKVDGKYYVYLKKA 119  
DB 62 TSHGDHYHYNGKVPYDALISELLMKDPNKLKDEDIVNKGKGYIKVDGKYYVYLKKA 121  
QY 120 AAHADNVRTEKEINROKQESHQHEGCTPRNDGAVALARSGRTTDDGYIFNASDIIEPT 179  
120 AAHADNVRTEKEINROKQESHQHEGCTPRNDGAVALARSGRTTDDGYIFNASDIIEPT 179  
DB 122 AAHADNVRTEKEINROKQESHQHEGCTPRNDGAVALARSGRTTDDGYIFNASDIIEPT 178  
QY 180 TGDAYIVPHGDHYHYPKNEISASELAEEAFISGRGNLSNRTYRRONSDNTSRTNWPVS 239  
180 TGDAYIVPHGDHYHYPKNEISASELAEEAFISGRGNLSNRTYRRONSDNTSRTNWPVS 239  
DB 179 TGDAYIVPHGDHYHYPKNEISASELAEEAFISGRGNLSNRTYRRONSDNTSRTNWPVS 227  
QY 240 SVSNPCTNTNTSNNSTNSQASQNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 299  
240 SVSNPCTNTNTSNNSTNSQASQNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 299  
DB 228 NPAQRLSENHNLVTPYTHQ-NOGENISLLRELIAKPLSRHVESDGLVFPDQAQITST 286  
QY 300 TARGAVAVPHGDHYHYPYQSELEERLARIPLRYRSNHVPPDSRPEOPSPOTPEPSP 359  
300 TARGAVAVPHGDHYHYPYQSELEERLARIPLRYRSNHVPPDSRPEOPSPOTPEPSP 359  
DB 287 TARGAVAVPHGDHYHYPYQSELEERLARIPLRYRSNHVPPDSRPEOPSPOTPEPSP 346  
QY 360 GOPAPNLK-IDSN--SSLVQAVRKVGEVVEEKGISRYVFAKDLPSFTVKNLESKLSK 415  
360 GOPAPNLK-IDSN--SSLVQAVRKVGEVVEEKGISRYVFAKDLPSFTVKNLESKLSK 415  
DB 347 SFOPAPNPQAPSNPIDEFLKAEYKAVKAGDGYVEENGYSRIIPAKDLSETAAGIDSKL 406  
QY 416 SKQESVSHTLTAKKENVAPROEFYDKAYNLTFAHKALEFNKGRNSFOALDKLELNDESTN 475  
416 SKQESVSHTLTAKKENVAPROEFYDKAYNLTFAHKALEFNKGRNSFOALDKLELNDESTN 475  
DB 407 AKQESLSHLGAKKTTDPSDREFFYNKAYDLARIHQDLDDKKGQVDEALDNLERLK 466  
QY 476 DESTNKEKLVDDLAFIAPITHPERLGRKNSQIEYTEDEVRILAOIADKTTSDGIIFENH 535  
476 DESTNKEKLVDDLAFIAPITHPERLGRKNSQIEYTEDEVRILAOIADKTTSDGIIFENH 535  
DB 467 DVXSDKVLVXDILAFIAPITHPERLGRKNSQIEYTEDEVRILAOIADKTTSDGIIFENH 526

OY	536	DIISGEGAAVTPHNGSHWIKGDSLSKKEVAAQATRKESGLPPSPADAVKANAPTGD	595
Db	527	DITSDGGAAYTPTHHTSHWIKDLSSEAKRAAQAIVAKESGLTPSTHDQSGNTAEKG	586
OY	596	AAAIYNRKGKERIPLVLYPWVEHTVEKKNNLLIPKHCHHNIRKFAMFDHTYKA	655
Db	587	AEAATNRKAAKRVPLDMRPVYLQYTVEKNSSLIIPHVDHNNIKFEWFEDGCLTEAPRG	646
OY	656	YTLDELFTATITYYEHPEERPHSNDGMGNASHVYTKCKDHSEDPKNKEADEE-----P	709
Db	647	YTLDELFTATITYYEHPEERPHSNDGMGNASHVYRNKGADTNQTEKPSEEEKQTDEK	706
OY	710	VEET-----PAEPTRPOVERETEYKAOLKEAEVLAKYTD	743
Db	707	EETPREKREKPSSEKPESEKPTPEEPSESEBESEHPVEYTEKEVKELREADDLGKI	763
 RESULT 3 US-09-765-272-182 Sequence 182; Application US/09765272 Patent No. US20020061545A1  GENERAL INFORMATION: APPLICANT: Choi et. al. TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines NUMBER OF SEQUENCES: 452 CORRESPONDENCE ADDRESS: ADDRESSEE: Human Genome Sciences, Inc. STREET: 9410 Key West Avenue CITY: Rockville STATE: Maryland COUNTRY: USA ZIP: 20850  COMPUTER READABLE FORM: MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage COMPUTER: HP Vectra 486/33 OPERATING SYSTEM: MSDOS version 6.2 SOFTWARE: ASCII Text CURRENT APPLICATION DATA: APPLICATION NUMBER: US/09/765,272 FILING DATE: 22-Jan-2001 CLASSIFICATION: <Unknown> PRIOR APPLICATION DATA: APPLICATION NUMBER: 08/961,083 FILING DATE: <Unknown> ATTORNEY/AGENT INFORMATION: NAME: Brookes, A. Anders REGISTRATION NUMBER: 36,373 REFERENCE/DOCKET NUMBER: PB340P2 TELECOMMUNICATION INFORMATION: TELEPHONE: (301) 309-8504 TELEFAX: (301) 309-8512 INFORMATION FOR SEQ ID NO: 182: SEQUENCE CHARACTERISTICS: LENGTH: 447 amino acids TYPE: amino acid STRANDEDNESS: single TOPOLOGY: linear MOLECULE TYPE: protein SEQUENCE DESCRIPTION: SEQ ID NO: 182:			
 US-09-765-272-182			
Query Match                    28.5%; Score 1187.5; DB 10; Length 447;			
Best Local Similarity 46.3%; Pred. No. 1.1e-70;			
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10.			
OY	4	LGLIQARTVKNENRVSIDGRQATDKTENLAPPDEVSKREGINAEOIVIKITDGYVTSHG	63
Db	1	LNHRSQENKCNRRNSYSYDGSSQSOKSENLPDPVOYSOKEGIQAEQIVIKITDGYVTSHG	60
OY	64	DHYHYNKRVYDDAITSELLMKDNRYLKDEDIVNEVKGGVIVKVDGKYVYYLKDADAHA	123
Db	61	DHYHYNKRVPLDALESSELLMKDNRYLQKADIVNEVKGGVITIKVDGKYVYYLKDADAHA	120

OY	124	DNRPRKEINNKOKHSHREGGTPRRNDGAVALARSGRYTTDDGYFENASDIIEPDGA	183
Db	121	DNVRKDKDEINNKOKEHVADNE-----KYSNVAAVARSOGRTTNDGYVFNADIIIEEDGNA	176
OY	184	YIVPHGDYHYIIPKNELSASELAAAEAFISGRGNLSMSRTYRRONSDNTSRTVMVSVSN	243
Db	177	YIVPHGCHYHYIIPRSDLSASLAAAKHLAGKNNGPQSLYSSTASDN-----	224
OY	244	PETTNTNNSNNSNNSQASNDIDSLIKOLYKLPPLSQRYVESDGLVFPDQAOTTSTARG	303
Db	225	-----NTQSVANGSSTSKPANKSENIQSLSLKELYDPSAQRYSSESGLVFDPDAKIISRPNQ	280
OY	304	VAVPHGDYHYIIPYKSOMSELEERLARIILPLRYSNHMVPPDSRPEOPSPOTPEPSPQOP	363
Db	281	VAIPIRGDHYHYIIPYKSLASLEEKIARRVP-----	309
OY	364	APNPLKIDSNSLSVOLYRKVGEGYVFEFGISRYVFAKDLPSETVKNLESKLSKOESVSH	423
Db	310	-----ISGTSGETVSTNAK-----PNEVVSLSGLSINPSS--	339
OY	424	TLTAKKEVAVPRDQEFYDVKAYNLLTEAHKALFXNKGKNSDFOALDKLERLNDSESTNKEK	483
Db	340	-LITSKELSSASDGYINPKP-DIYEELATAIYVHG--DHFHYIYK-----SNQIGOPT	389
OY	484	LVDDLAFIADITPHERLGPKNQSIETYETEDEVRIAOIADLADRYTSDGYIFDEHDIISDGD	543
Db	390	LPNNSLAPPS-SLPINPGTSHEKHE-----EDGYGFANRLIAEDES	431
OY	544	AYVTPPHMGHSHWIK	558
Db	432	GFVMSHGDNHNYFFK	446

```

RESULT 4
US-09-815-242-5816
? Sequence 5816, Application US/09815242
? Patent No. US20020061569A1
? GENERAL INFORMATION:
? APPLICANT: Haselbeck, Robert
? APPLICANT: Ohlsen, Kari L.
? APPLICANT: zyskind, Judith W.
? APPLICANT: Wall, Daniel
? APPLICANT: Tiawick, John D.
? APPLICANT: Carr, Grant J.
? APPLICANT: Yamamoto, Robert T.
? APPLICANT: Xu, H. Howard
? TITLE OF INVENTION: Identification of Essential Genes in
? TITLE OF INVENTION: Prokaryotes
? FILE REFERENCE: ELITRA.011A
? CURRENT APPLICATION NUMBER: US/09/815.242
? CURRENT FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: 60/191,078
? PRIOR FILING DATE: 2000-03-21
? PRIOR APPLICATION NUMBER: 60/206,848
? PRIOR FILING DATE: 2000-05-23
? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 5816
? LENGTH: 2478
? TYPE: PRT
? ORGANISM: Staphylococcus aureus
US-09-815-242-5816

```

Query Match 4.1%; Score 170; DB 10; Length 2478;  
Best Local Similarity 19.4%; Pred. No. 0.0072;  
Matches 169; Conservative 141; Mismatches 354; Indels 206; Gaps 38;

```
QY 8 QARTYKNNRVSYIDGKQATQKT-----ENLTPEVSKREGINAEOIYIKITDGGVYS 61
DB 986 QATTTKSKAKKEIQAOKASERKTAIEAMNDSTTEEQAAKDKV-----DQAVVTA 1035
QY 62 HGD-----HHYHNGKVPYDAIISBELMKDPNKKLDEDIYNEVKGGVYKIVDGRYYV 115
DB 1036 NADIDNAANNVDNAKKTNEATTA--ATTPANVYPAKQAIADKVQAQETAIDG----- 1089
QY 116 YLKDAAHADNVPTKEEINRQKOEHSOHRREGTPRNDGA-----VALAR---SOG 161
DB 1090 -----NNGSTTEEKAAKQOVOTEKTTADAIDAHTNAEVEAKKAIAIKETAIOP 1141
QY 162 RYTTDDGYTFNMSDIIETDGAIVPHGDHYHYPKNEISASELAEEAFLSGRNLSNS 221
DB 1142 ATTTKD-----NKEAATATANKERKTA-----IAOTODITAEIAAANA-----DVDNA 1185
QY 222 RYTRRONSNTSRNTNVPVSNGTNTNTSN--NSNTNSQASQSDIDSLKQLYKLP 279
DB 1186 VT---QANSNIEANSQNDVDAKTTGENSIDOVTPTVKKATARNETAIILNN--KLQE 1240
QY 280 SQRHVESDGLVDPDPAQITSRARGVAVPHGDHYHYPYSQMSLEERIRARIITPLRYSNH 339
DB 1241 IQATPRAIDEEKQAADAENKENGKA-----NQAISAAITTAQVDEAKNANEA 1288
QY 340 WYPSRPEQSPQPTPEPSPGQPA-----PNLKIDNSLSVQLYRKVGEG----- 386
DB 1289 AINAVTPKVVKKQAQKDEIDLOAQOTQNTVINNDQNAATEEKEAALIQOLATAVYDAKNNIT 1348
QY 387 YVEEKGISRYVFAKDLPESEYVKNLES-----KLSQESVSHTLTAKKENVAPRDGEFYDK 442
DB 1349 AATDDNGVDO---AKDAGKNSIOSTOPATAVKSNAKNVDQAVTTQNOAI-----DN 1397
QY 443 AYNLLTEAHKALFXNKNRSDFOALDKLLER--LNDESTN-----KEKLYDDLAFAPIT 496
DB 1398 TTGATTEE-----KNAKADLVILAKAKAYODILMAQTNDVYQIKQAVADIOGITADTT 1452
QY 497 HEBRLCKPMSQIETTEDVRIQAOLADKTTSDGYTFDEHDIITSDCEDAVYTPHMGSHMT 556
DB 1453 IAD-VAKDELATRKANQOKALIQOTADATT-----EKEEONNOQVADQOLT--QGNQNTIE 1502
QY 557 GKDSLSDKXVAQAAYTKKGIILPSPDADYKANPTGDSAAIYVNRKGEKRIPIVRLRY 616
DB 1503 NQOSIDVDTAADMAL---QALDPIDASTDVKTNNARAELLTEKONKT-----TEILN 1551
QY 617 WEHTVEVKNGLIIP---HKDHYHNKFAWF--DDHTYKAPNGYTLLEDPATIKYYVE 670
DB 1552 NNHTTEER--GNDIGFVRAVEGLNINNAATTTGAVTAKDTAVQKVOQLHA----- 1603
QY 671 HEDERHSDNGMNSAEHLGKKHSDHEDPNKNFKADEE--PVEETP--APEEVPQVETEYK 727
DB 1604 NPVKRP-----AGKKEILDQ-----AAADKKQIOTIOTPAWSOEEINDAKOE--V 1644
QY 728 EQOLKEA-----EVLAKVYDSSILKANATETLAGLRNNLTLOIWD----- 767
DB 1645 DIELNDAKTNNDOSSITNEYDNAVKEGKAKINAVKTFSEYKDALAKIEDAVINAKYNEAD 1704
QY 768 -----NNSIMAEKILALLKGSNPSVSK 792
DB 1705 NSMASTSSIEAEKKOELAKOTADQNVNQ 1734
```

RESULT 5  
US-09-815-242-12967  
Sequence 12967, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.

```
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ElitRA-011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 12967  
LENGTH: 2478  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12967
```

Query Match 4.1%; Score 170; DB 10; Length 2478;  
Best Local Similarity 19.4%; Pred. No. 0.0072;  
Matches 169; Conservative 141; Mismatches 354; Indels 206; Gaps 38;

```
QY 8 QARTYKNNRVSYIDGKQATQKT-----ENLTPEVSKREGINAEOIYIKITDGGVYS 61
DB 986 QATTTKSKAKKEIQAOKASERKTAIEAMNDSTTEEQAAKDKV-----DQAVVTA 1035
QY 62 HGD-----HHYHNGKVPYDAIISBELMKDPNKKLDEDIYNEVKGGVYKIVDGRYYV 115
DB 1036 NADIDNAANNVDNAKKTNEATTA--ATTPANVYPAKQAIADKVQAQETAIDG----- 1089
QY 116 YLKDAAHADNVPTKEEINRQKOEHSOHRREGTPRNDGA-----VALAR---SOG 161
DB 1090 -----NNGSTTEEKAAKQOVOTEKTTADAIDAHTNAEVEAKKAIAIKETAIOP 1141
QY 162 RYTTDDGYTFNMSDIIETDGAIVPHGDHYHYPKNEISASELAEEAFLSGRNLSNS 221
DB 1142 ATTTKD-----NKEAATATANKERKTA-----IAOTODITAEIAAANA-----DVDNA 1185
QY 222 RYTRRONSNTSRNTNVPVSNGTNTNTSN--NSNTNSQASQSDIDSLKQLYKLP 279
DB 1186 VT---QANSNIEANSQNDVDAKTTGENSIDOVTPTVKKATARNETAIILNN--KLQE 1240
QY 280 SQRHVESDGLVDPDPAQITSRARGVAVPHGDHYHYPYSQMSLEERIRARIITPLRYSNH 339
DB 1241 IQATPRAIDEEKQAADAENKENGKA-----NQAISAAITTAQVDEAKNANEA 1288
QY 340 WYPSRPEQSPQPTPEPSPGQPA-----PNLKIDNSLSVQLYRKVGEG----- 386
DB 1289 AINAVTPKVVKKQAQKDEIDLOAQOTQNTVINNDQNAATEEKEAALIQOLATAVYDAKNNIT 1348
QY 387 YVEEKGISRYVFAKDLPESEYVKNLES-----KLSQESVSHTLTAKKENVAPRDGEFYDK 442
DB 1349 AATDDNGVDO---AKDAGKNSIOSTOPATAVKSNAKNVDQAVTTQNOAI-----DN 1397
QY 443 AYNLLTEAHKALFXNKNRSDFOALDKLLER--LNDESTN-----KEKLYDDLAFAPIT 496
DB 1398 TTGATTEE-----KNAKADLVILAKAKAYODILMAQTNDVYQIKQAVADIOGITADTT 1452
```

0y	497	HPBELYGPNQSIETDEEVNIAQALAKYTTSDGJFDEDDIISDEGNAVYPRHMOHSWI	556
Db	1453	IKD--VAADELATYKANEOKALIACTAATTT-----EEREQANQOVDKOLT--QGNOMIE	15020
0y	557	GKDSLDSKEVYAAQATYKKEGILPSPDADYKANPQGSAAIYVRVNGEKRIPLVRPY	616
Db	1503	NAGSIDDVNTAKDNAT---QALDPIDASTDVKTNAEALITTEMOKTI-----TELLN	15510
0y	617	MVEHTYEVKNGNLIIP---HKDHYHNIFEAF--DDHTYKAPNGYTTLEDLFATIKYVE	670
Db	1552	NNEFTTNEEK--GNDIGPRAVEEGELNNINAAPTTGGVTTAKOTFAVQKYOQLHA-----	16030
0y	671	HPDERPISNDGMCWGNASEHVLGKDKHSDPEPNKFNKADEE--PVETP--AEPEVPOVETEV	727
Db	1604	NPVKKP-----AGKKELQDQ---AAADKKTQIOTFPNASQOEITDAKQE--V	16440
0y	728	EAOJKEA-----EVLAKVYTOSLSKANATETLAKRNLLQJLMD-----	767
Db	1645	DTELNQAKTINVDSSTINEYVDNAVKKGAKINAVKTFSEYKAKDALAKIEDAYNAKVNAD	17040
0y	768	-----NNSIMAEAEKILLALLKGSPSPSVAK	792
Db	1705	NSNASTSSEIAENAKOKALDELKOTAPDANVMO	1734

```

: RESULT 6
: US-09-946-805-4
: Sequence 4, Application US/09946805
: Patent No. US20020116734A1
: GENERAL INFORMATION:
: APPLICANT: Dickman, Martin B.
: TITLE OF INVENTION: PLANT DERIVED BAG HOMOLOGUES
: FILE REFERENCE: 480140.469
: CURRENT APPLICATION NUMBER: US/09/946,805
: CURRENT FILING DATE: 2001-09-04
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 1043
: TYPE: PRN
: ORGANISM: Arabidopsis thaliana
: US-09-946-805-4

```

Query Match	4.0%;	Score 168.5;	DB 10;	Length 1043;
Best Local Similarity	18.5%;	Pred. No. 0.0026;		
Matches 169;	Conservative 108;	Mismatches 327;	Indels 311;	Gaps 38;

```

0Y      40 KREGNMAQOIKTITDGGVYSHSDHHHYNGKV---PYALISEELMMDPNKLKDED   96
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      271 KRDDVEASE-----SSNDKRCKMOGKVVEYPEDISMISL-               308
0Y      97 IVNEVGKGVIKVDGKYVVYLKDAHAHADNVNTKEEI-----NRQOEHSQH   142
       ::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      307 ---IQG-----QDVKEAQNOQNKKEPEQVPPIPIWIPSYGRKRVSES         348
0Y      143 REGGP-----RNDGAVALARSOGRYTTDGCIYFNASDIETDGDAYIVPHCD   190
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      349 KESSEGNHLESCPSDLHRNEGQTQAKGK-----EONFECNVLSDAEKSVINIPVAN   403
0Y      191 HY-----HYIPKNELASSELAAEAFLSGR-----                215
       ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      404 HLOEPRIAPVKLSENHLPKPIEPPTKRIAKNPVKSTYKKEOSSSSESKLPPLCLRVDP   463
0Y      216 ----GNLSNRTYRRONS DNT-----SRTNWVPYSVNPGTINTNTS---NN    254
       |::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      464 PKERNCGKSVSHPKRMREKSKETKIAPLSKKAESFT--VPEACYNKCEDANAEMKMAE   521
0Y      255 SNTNSQAQSANDIDSLKQLTKLPLSQORHNVESOLVDPDAQITSRTARGAVAPHGHYHF   314
       ::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      522 GSLNALMRKEKGSVES-----NSNLQDESNELIITPCPEAKENRBP-----        561
0Y      315 IPYSMSELERIRIIRIIPLRYS---NHWWVDSRPDEQSPQPOTPEPSPG-----        360
       ::||::||::||::||::||::||::||::||::||::||::||::||::||

```

```

Db 562 ---AKKSTTEEPARIIGSMYRGCDVNRMRPIKKLEIA---TVBQMDXCKKRILEALEA 615
Qy 361 --POPAPMLKIDNSNSSLVSQLYKRVGEGVYFEEEKGISRYVFAKDLPSFTVKNLSEKLS-- 416
Db 616 STDQHEEKEETIVNGELVMNLLIKLD---AVEGLHPSIREFKALATE--LSSIQOKLSD 671
Qy 417 KOESVSHTLTKAKKNVAPPROGEYDRAYN---LLEAKKALEFXNKGKRSNDQALDKLLE 472
Db 672 KNSCASAEKEVVKQOYEIKQOP--SDSPVNLHSQLTEENKMV----- 712
Qy 473 RLNDSESTKEKELVLDLAPITTHPERLQKPNQSOLEYNEDEVRIALQADKYTTSDGYIF 532
Db 713 ----SDTNLEAVLR-----LSPEEH-----PMSVNLRIDEX---QAESAEETEGY-- 751
Qy 533 DEHDIIISDEGDVAVYTPMGHSHWIGKDSLSDKKEKVAQAAYT-----KEKGILPSPDA 585
Db 752 ----GFLTLDATDSKQATENMAAASSTIPEKIGEVETVVGNP-- 791
Qy 586 DYKANPFGDSAAAIYNRVKEEKRI-----PLVRLPYVEHT-----VEV-- 624
Db 792 ----PSADNGMGTVTVEENKAMAVSELEEPINELPQWVEETETNSIRDPENASEVSEA 846
Qy 625 ----KNGLLIIPHD---HHHNIKFAFDHT---YKANPGTILEDLFA--- 663
Db 847 ETNSSENERRRGEDDVLHSEKEVSEISELPVGVIDETQPLSDQSPSSYTRREGNMTAMP 906
Qy 664 -TIRYVEHPERDHSNDGNGNASEHVLGKDDSEDPNKNFKADEEYPV-----ETPAE 716
Db 907 KTAQOELEVDHDSNNNSKGGIGOOTSEPODEKES--PEFEVLYKQEPLETEVYLMEQAPE 964
Qy 717 PEVPOVETEKVEAOLKEAEVLLAKVYDSSLKANATETTLAGLRNNLTLOIMDNNSIMAEAE 776
Db 965 PEITEPQISKEETKKLMBENQRFKMETETLVKAGREQ--LEVISKILTSRYKSLLEKLSHK 1022
Qy 777 KLALALKGNSPSSYS 791
Db 1023 KTOIRRRASKEMVS 1037

```

```

RESULT 7
; US-09-815-242-5493
; Sequence 5493, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyckind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELTRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/244,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,635
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

```

SEQ ID NO 5493  
LENGTH: 870  
TYPE: PRF  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5493

Query Match 3.8%; Score 159; DB 10; Length 870;  
Best Local Similarity 19.3%; Pred. No. 0.0086;  
Matches 164; Conservative 120; Mismatches 273; Indels 294; Gaps 45;

14 ENNRVSYIDGKATOKTEMLT-----PDEVSKREGIN-AEQIV 50  
DB 54 QNNNATTTTTPPNDOTNOTOPATOPANTAKNYPAADESLKAIKNPAVENKEHDIQPREQVN 113  
QY 51 IKITDOGYVTHSGDHYHNGKVPYDAIISELLMKDPNYKKLDEDIVNEVKGVIKYD 110  
DB 114 FOLLDKNNET---QYHFFSIKDPADV---YTKKAEVELDINTASTWK-- 157  
QY 111 GRYYYLAKDAHADNVRTKEEINRCK-----QESOHREGGTPRNDGAVALAR 158  
DB 158 -KFEVY-----ENNOKLPVRLVSYSPVEDHAYIR--FPVSDGTQEL-K 197  
QY 159 SGGRTYTDGY-----IFNASDILE-DTGDATVPHGDHYHYPKNELSAS 203  
DB 198 IVSSQIDGGAETNDYTKLVFAKPIYNDPSLVKSDTDNA-----VATNDQSSS 246  
QY 204 ELAAAEAFLSGKGNLSNSTRYRONSNDTSRTNWVP---SVSNPG----- 245  
DB 247 DAS-----NOTNTNT-SNONTSTNNANNQPOATDMSQPAQKSSANADQASSQ 295  
QY 246 ---TNTNTSNNSNTNSQASQSDIDSLKOLKPLSLGRHVESDGLVD---PAQITS 298  
DB 296 PAHETNSNGNTNDKTNESSQSD-----VNOQYPPADESLQDAIKNPAIIDK 342  
QY 299 RTARGAVALPHGDHYHYPYSQMSLEERARIIPLRYSNMHWPPDRPSPQPPPS 358  
DB 343 EHT-----ADNWRPIDQMKNDGER-----QYHYHA-----STVEPA 375  
QY 359 -----PGQAPAPNLK-----IDNSSLSVQLVRKVG---GYV-FEEKGISRYV 398  
DB 376 TVIFTKTGPIELGLKTAJSTWKKEFEYEGDKLVPVELSYSDSKDYAIRFPVSNGTREV 435  
QY 399 FAKDLSEVYKULESKLSQESVSHLTAKKENVAPROEPFD-KAYN---LTTEAHKAL 454  
DB 436 -----KIVSIEYGENIHEDYDTLWFAOPTNNPDYDEETYNQOKLLAPYHKA- 487  
QY 455 FANNGNSDFQALDKLLERLNDE-STNKEKLVDDLLAFIA-----PITPERLGRKNSOI 508  
DB 488 ---KTLERQVYELKLEKLEKPEKRYKAKKIDQTVRELADQVKSAVTEFENVTPINDQ 544  
QY 509 EYTEDVIRIAQLADKYTTS--DGYTFDEHDIISD--EGDAYVT-PRMGSHWIGKDSLSD 563  
DB 545 TVQGEAHFVFESEENSESVMDGFV--BHPFYATLNGQKYVMKTKDSDYM--KDLIVE 600  
QY 564 KKKVAAQAF-TKEG---LLPSPDADVANKPTGDSAAAIYNRVGEKRIPLVRLPYME 619  
DB 601 GRVTVTSKDTLNNSKTLFPYIPD-----KAVYNAI--YKVVANANGYESQ 645  
QY 620 HTEVEVANGMLIIPHKDHYHNIKFAWFDHTYKAPNGYTTLEDIFATIKYVEHDERPHSN 679  
DB 646 YAVRIIN-----QDIKTQ--DDDTSQ----- 664  
QY 680 DQMGNASEHYLCKKDHSEDPKNKFKADEEPVEETPAEPVPOVETKEVEAQKKEAVLLA 739  
DB 665 ---NNTSEPLNQTQGOEDNVATDTATNNSIETPSE-----ADDKVD--LIEPESDMV 713  
QY 740 KYTDSLSKANA 750  
DB 714 KQTDSSVDKDA 724

Sequence 12637, Application US/09815242  
Patent No. US2002061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815.242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12637  
LENGTH: 870  
TYPE: PRF  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12637  
Query Match 3.8%; Score 159; DB 10; Length 870;  
Best Local Similarity 19.3%; Pred. No. 0.0086;  
Matches 164; Conservative 120; Mismatches 273; Indels 294; Gaps 45;





```

: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 12610
: LENGTH: 5795
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
: US-09-815-242-12610

```

```

Query Match          3.6%; Score 151.5; DB 10; Length 5795;
Best Local Similarity 18.5%; Pred. No. 0.39; Indels 253; Gaps 40;
Matches 174; Conservative 133; Mismatches 381;

```

```

OY 1 SYELGLYQARTYKEN-----NRVYIDGKOATQKTEMLTPPEVSKREGI-----NA 46
DB 2680 SOATNLAAVESKQSANSLDGMNGNQTALINDKSGTASQNLDELDECKRNNVNAQVSA 2739
OY 47 EGVIVKIDGQVYTHSGDHYHYNGKVPYDAIISELMLKDPN---YKLKDEDIYNEVGK 103
DB 2740 ETLNKQKTS-----PNTAKTAVEQALNNVNSAKHALNGTQNNNAKQ 2781
OY 104 GVIYVDG-----KYVYVYKDDAAH-ADNVPRKEELNRKQKSHSOHREGGTPRN---DGA 153
DB 2782 AAITALNGASDLNOKOKALKAQANGAQRVSAQVORNRATE--LNTAMGTILKHAIDKT 2839
OY 154 VALARS-----QGRYTTDDGYIFNADLIEDTGDAYIVPHGDHYHYIFKNELASAE 204
DB 2840 NTLASSKYVNAADSTKQNAVTTK---VTNAEHITISGPIVVTTP-----SE 2881
OY 205 L-AAEAFLISGNGNLSNSETTYR--RONSNDNISRTKMWPSVSNPGTTNTSNNSTNSQA 261
DB 2882 VTAAANQVNASAKOELNGERLFRVAKQA-----NTAIDATQLTLPQAKLKEQV 2931
OY 266 SOSNDIDSLKQLYKLPISORHVESDGLVDFPAQ---ITSRTARGAAYVPHGDHYHYHPIYSQ 319
DB 2932 GQANLELDI-----QTVOTNGQALNNAMKGLRDSINETTYVASQYNTDASPNN 2980
OY 320 MSELBERTAIIPLRYRSNHWVDSRPEQSPQPTPEPSPQAPNLKIDSNS-----373
DB 2981 OSTYSAVSNMAGIINTQNTPTMDTSAITQATQYVNNAKNGLNGAENLRNAQNTAKONIN 3040
OY 374 -----SLVSQLVKRVGEGVYFEENGISRYVFAKDPSE---TYKNLESKLSKQES 420
DB 3041 TLSHLTNNOKSAISSOIBRAGH-----VSEVTAAKNAATELMTQMGNTLOALHDQNT 3092
OY 421 VSHITLAKKENAPRDOEYDKAY-NLITLAKALFXNKGKRSDDQALDKLLER-----473
DB 3093 VKQGNFTDADAKRD-----AYTNVARSRAETILNKTOGANTSQDVEAIAIONVTSAKN 3146

```

```

OY 474 -LNDES--TNKEKYVDLLAFLAPITTPERLQKPN-----SOLEYTEDEVRIAQ 519
DB 3147 ALNGQONTNNAKNTAKHALNMLTSSINNAQKRDLTTRKIDQATVAGVBAVSNGTQQLNAM 3206
OY 520 -----LADKYTT--SDGIYFDEHDIISDEGDAYVTPPHGSHWIKGKSLSKREYAAO- 570
DB 3207 ANLQNGINDKANTVLASENY-----HDASDKRTAYQAVTNAENIINKSGSNLDRKAIVEN 3262
OY 571 ---AATKKEGII-PPS---PDADVKNAPDGSAAALYNNRKSGEKRIPLYVLPVVEHTVAV 624
DB 3263 ALSQVTNNAKALNGHNLNLEQAKSNANTT-----INGLOHLTTAOKDKLKQOVQQA 3312
OY 625 KNGNLIIPKRDHYHNKIFAMFDDHTYKAPNGYTLDELFAITIKYYE-----HPDE 674
DB 3313 QN-----VAGVD--TVKS-SANTLNGAMGTLRNSIQDNTATNNGQNYLDA 3354
OY 675 RPHSDNGMGNASEHVLGKRDHSEDPKKNKA-----DEEPEETPAPPEVQVET 724
DB 3355 TESNKTNNNAVDASANGVYNATSNPMDANAINOIATQVTSKMLDGTHTNLTOAKQAT 3414
OY 725 EKVEAQUR-----EAEYILAKVTDSILKANATEFLAGLRNLT-----LO--INDNNS--- 770
DB 3415 NALDQATNLNKAQKALKAQVTSQORVANVT--SIQOTANELTAMGQLQHDGIDDERATQ 3473
OY 771 -----IMAEAEKLALILKSGNPSVSKER 794
DB 3474 TQKRYRDAQSKKTAYDQAVAAKAILNKQTSNSDKAAVDR 3514

```

```

RESULT 11
US-09-815-242-5635
: Sequence 5635, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:

```

```

: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 5635
: LENGTH: 2368
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
: US-09-815-242-5635

```

```

Query Match          3.5%; Score 145.5; DB 10; Length 2368;
Best Local Similarity 18.2%; Pred. No. 0.27; Indels 295; Gaps 34;
Matches 156; Conservative 122; Mismatches 285;

```



Db 1396 AKDNAL---QALDPIQASTVDVKTNAFALLTEMQNKI-----TELLNNNETTEEK- 1443  
QY 627 GNLIIID-----HKDHYNIKFAWFDHDTYKAPNGYLTLEDLFAITKYVHEHPDERPHSDQM 682  
Db 1444 GNDIGCVRAAYEEGLNNT-----NBAITTGVTAKDTAVOKVQOOLHNP-- 1488  
QY 683 GNASEHVLGKDKHSDPNKFNKRADEE-PVEETP--AEPEVPOVETEKVEAQLKEAEVLLA 739  
Db 1489 -----VKRPAKGTALDQAADKKTQLEQPPNASSQOEINAKOE-VDTELNOAKTIND 1539  
QY 740 K-----VTDSSLKANAFETTLAGLRNNLTLOI-----MDN-----NSIMAE 774  
Db 1540 QSTDEYVDNAVKEGKAKINAVKTFSEYKRDALAKIEAAYNAKVTEADNSNASTSEIAE 1599  
QY 775 AEKILALLKGNPSVSK 792  
Db 1600 AKOKLAELKOTADQNVNQ 1617  
RESULT 13  
US-09-765-272-118  
Sequence 118, Application US/09765272  
Patent No. US20020061545A1  
GENERAL INFORMATION:  
APPLICANT: Choi et. al.  
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines  
NUMBER OF SEQUENCES: 452  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/765,272  
FILING DATE: 22-Jan-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/961,083  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brookes, A. Anders  
REGISTRATION NUMBER: 36,373  
REFERENCE/DOCKET NUMBER: PB340P2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1040 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 118:  
US-09-765-272-118  
Query Match 3.5%; Score 144; DB 10; Length 1040;  
Best Local Similarity 19.2%; Pred. No. 0.11;  
Matches 154; Conservative 111; Mismatches 284; Indels 252; Gaps 38;  
QY 107 IKVDGKYV-YLKDAHADNVRTKEINRQKQHSQHREGGTPRND-----G 152  
Db 76 LKIEGQYIGYIK-TKKQONTLSRVDKYSQNRQSNSTKTSVDVHASADLEWQGG 134  
QY 153 AVALARSQGRYTTDGGIIFNASDIIEDTGDAYIVPHGDHYHYIPKNELASASLAAEAFL 212

Db 135 KVSLL---QGEASGDG-----LSEKSSIAADNLSSNDSPA 166  
QY 213 SGRGNLSNRRTYRRONSDTSRTNVPSSVNGPTNTNTSNNNSQASQNSDLSLK 272  
Db 167 S-----QVEQNPDKHGESYVRPTYPEQGNVYSATTVQSABEEVLATIND-----R 211  
QY 273 QLYKLPLSGRHVESDG-----LVF-----DPAQITS 298  
Db 212 PEYKPLFETKGTQEPGHEGEAAVRDLPLYTKPLFETKGTQPGGHEGEAAVRPEEPATYEP 271  
QY 299 RTAGVAVPHGHHYHPIPSQMSLEBERAKIIPLRYSNMHNVPPSRPPSPQPTPERS 358  
Db 272 LATKGTQEPGHE-----GKATVREETLEYTEPAVATGT-----QEPHEGERVVEEL 319  
QY 359 PGCPQA-----PNT-----KIDSNSLVSQVLR---KVEGCVFEE 391  
Db 320 PALEYTRNRKTELQNIPTYTEIOPDTLLKNRKRTERQQAQTRITQYEDYIVNGNVET 379  
QY 392 KGISRYVFAKDLPSFTVKNLESLSKQESVSHTLAKKENVAPRQDEYDAKYNLL--TE 449  
Db 380 KEVSRTEVA--PVNEVYVGTIVKVPVEITNLTKVEN---KKSITVSYNLTDTTS 431  
QY 450 AHKALEFXNKGSRSDQALDKLERLNDESTNKEKLVDDLAFARI--TH-PERLGRPN 505  
Db 432 AYVS-----AKTQVPHG--DKLVKEVDIENPAKEQVYISG-LDYTTYTYVKTULTYMLGENN 484  
QY 506 SLEYTEDEVRJAOLADKRTTSDGYTFDEHDIISDEGDAYVPHMGHSHWIKSDLSKE 565  
Db 485 E--ENTETSTQFPOL--EKKEIKDIDSEVLXGENDRY-----RXYL---SLESPR 530  
QY 566 KVAQAAYTKKGIILPPSPDADYKANFTGDSAAIYNRVKGERRIPLVRLPYVEHT--- 621  
Db 531 TDTAKYFVKV-----SDRFKEMYLVPVS-----ITENTDGY 563  
QY 622 -VEVKNGLIIPHKDHYNIKFAWFDHDTY-----KAPNGYT-LEDLFAIK-----Y 667  
Db 564 KYTVAVDQLEVEGTGQYK-----DDYTFYVAKSKAEPGYTFSKQLVYMAQNSLGV 616  
QY 668 YVEHDERPHSHDGMGNASEHYLAKKQHS-----EDPKNFKRADE--EPVEET 713  
Db 617 YTLASDM-----TADSVSLGDKQTSYLTGATGSLGSDGKRSATYILKKRPLDPT 667  
QY 714 PAPEVPOVETEKVEAQLKEAEVLLAKYTDSSL-----KANATETLAGLRNNLTLOI 765  
Db 668 LMGATVRRDDITVTSADSKEVNAALAKAANSANINNVAVEKISGAKSVAGVAGSATVTV 727  
QY 766 MDNNSIMAEAEKLLALKGSN 786  
Db 728 IENSSFTG---KLIANHODSN 745  
RESULT 14  
US-09-815-242-5703  
Sequence 5703, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: Elittra.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848



Db 1542 YQAAVQ-----HAKDLINQTSNPTL-----DKAQVEQLT 1570  
OY 213 SG-----RGNLSNSRTYBRQNSDNTSRTNVPSPVNGTNTNUS--NNSNTNSQASQ---- 263  
Db 1571 QGVNNAKNDLHGDQKADDDQKAVTDLNLQNSLNPNQRALESQIINNAATREDEVAQKLA 1630  
OY 264 SMDIDSLKQLYKLPISQRHVESDGLVEDPAQITSTARGAVPHGDHY-----HFIPY 317  
Db 1631 AQALDQAMQALRNSIQDQOQTES--GSKFINEDKPOKDAVQAQAAVQAKDLINQTSNPTLDK 1689  
OY 318 SQMSLEERIAIIPURYSNNHWPDSPREQSPQPTPEPSPGQAPAPNLKIDS--NSSL 375  
Db 1690 SQVEQLTQAVTTAKNDLHGDQKADDDQAVTTVALPMLNHAQOQALTDALINAAPTRE 1749  
OY 376 VQOLVRKVEGVPEEKGISRYVFAKDLPEFTVKNLESKLS--KQESVHTLTAKENV 432  
Db 1750 VAQHQTATE-----LDHMETLKNKVQOVNTDKAQPNTTEASTDKKEAV 1794  
OY 433 APRDOEYDKAYNLLTEAHKALFXNKGNSDPQALDKLERLND-----ESTNKEKL 484  
Db 1795 ---DQAL--QAESITDDPTNG--SNANKDAVEQALTKLOEKYNELNGNERVAEAKAQAKQ 1847  
OY 485 VDDLAFLAPITHPERLGRPNQIETEDDEVRIQDLADKTYTSDGIIFDEHDIISDEGDA 544  
Db 1848 TTDOLAHL---NADQIATAKONIDQATKLOPIAEIYDQATOLNOSMDOLQOAVNEHANV 1903  
OY 545 YVTPHMGSHWIGKDSLSDEKEVAQAAYTKEKGIIPSPDADVKANPTGDSA-AAIYN-- 601  
Db 1904 EOTVDTYQAD-----SDKQNAVQALIAEAENVL-----KONSNKQOVDQALQNTL 1948  
OY 602 ---RVKGEKRIPLVRLPYWVEHTEVKNGNLIIPHKDHNIKPAWEDDHTYKAPNGYT 657  
Db 1949 NAKQALNGDEVALAK-----TNGKHDI--DQNLNLNNAQOQGFGRIDQSHD 1994  
OY 658 LEDLPAITIKYYEHDPDERHSDGNGNSEHYLGKKHSEDPKNKPKADEEPVEETPAEP 717  
Db 1995 LNOIQOIV-----DEAKALNRAMQOLSQETISGNEGRTKGSTNYVNAD-----T 2037  
OY 718 EYPOVETEKEVAQLKEAEVYLLAKYTDSSILKA-----NATETLAQLRNN 760  
Db 2038 QVKQYDEAVD---KAKQALDKSTGQNLTAQOYIKLNDVATKAKKALNGEERLNN-RKS 2092  
OY 761 LTLQIMDNNSIMAEAEKLLALIKSGSNPSSYSK 792  
Db 2093 EALQRIQDQTLHLNNAQROLAIQOINNAETLNK 2124

Search completed: November 13, 2002, 04:12:37  
Job time : 63 secs